

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 18:30:40 ; Search time 1981.17 Seconds
(without alignments)
95.427 Million cell updates/sec

Title: US-09-675-650-4

Perfect score: 20
Sequence: 1 gagtagagaagataagaacg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1022815 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	16.8	84.0	405	123	AW982117 SSS0027 S
C 2	16.8	84.0	411	141	BE859183 SSS0038 S
C 3	16.8	84.0	511	140	BE810166 SSS038 S
C 4	16.8	84.0	246	146	AZ609094 1M0433L11
C 5	16.8	84.0	619	104	AJ395794 AJ395794
C 6	16.4	82.0	196	167	BE448549 UTX2D01.Y
C 7	16.4	82.0	207	114	AW311691 NXMV055F1
C 8	16.4	82.0	231	138	BE648384 UT-M-BH2.
C 9	16.4	82.0	261	30	AV368821 AV368821
C 10	16.4	82.0	266	125	BB104945 BB104945
C 11	16.4	82.0	267	27	AV265878 AV265878
C 12	16.4	82.0	279	111	AW124239 UT-M-BH2.
C 13	16.4	82.0	287	125	BB100422 BB100422
C 14	16.4	82.0	292	128	BB217229 BB217229
C 15	16.4	82.0	293	127	BB155980 BB155980
C 16	16.4	82.0	307	133	BB426512 BB426512
C 17	16.4	82.0	308	164	BE155459 PM1-FH035
C 18	16.4	82.0	343	238	AZ078760 RPTC1-23-4
C 19	16.4	82.0	355	120	AW767597 d84b01.Y
C 20	16.4	82.0	366	114	AW311692 NXMV055F1
C 21	16.4	82.0	372	8	AA517758 AA517758
C 22	16.4	82.0	376	23	A1698906 wa81g08.x
C 23	16.4	82.0	396	8	AA499831 V905603.r
C 24	16.4	82.0	406	15	AI030547 UT-R-CO-3
C 25	16.4	82.0	409	138	BE688775 uw38e03.x
C 26	16.4	82.0	422	173	BE065373 H303D001-1
C 27	16.4	82.0	473	138	BE630423 ua20g03.x
C 28	16.4	82.0	485	18	AI315397 u46b08.y
C 29	16.4	82.0	486	14	AA254110 va08g07.r
C 30	16.4	82.0	508	18	AT134293 u124c05.x
C 31	16.4	82.0	509	174	BE141565 fa81b01.x
C 32	16.4	82.0	522	123	AW986420 u61g10.x
C 33	16.4	82.0	558	173	BE094146 ua72b07.x
C 34	16.4	82.0	568	139	BE753306 206275 MA
C 35	16.4	82.0	580	223	AQ001918 CIR-HSP-2
C 36	16.4	82.0	584	225	AQ159551 nbx00140
C 37	16.4	82.0	588	24	AT785259 u138c08.x
C 38	16.4	82.0	623	117	AW537380 G0115610-
C 39	16.4	82.0	647	243	AZ433919 1M0220J03
C 40	16.4	82.0	649	233	AQ780056 HS_3169_A
C 41	16.4	82.0	709	243	AZ409265 1M0180M11
C 42	16.4	82.0	849	170	BE866137 963066G12
C 43	16.4	82.0	1059	150	BE576271 602134137
C 44	16.4	82.0	1711	192	AK007734 MUS nmucu
C 45	16.4	80.0	458	225	AQ201343 RPTC11-43

ALIGNMENTS

RESULT 1
 LOCUS AW982117/c
 DEFINITION SSS0027 Snaeda salsa ZAP cDNA library Snaeda maritima subsp. salsa
 CDNA similar to putative nematode-resistance protein, mRNA
 sequence.
 ACCESSION AW982117
 VERSION AW982117.1 GI:8173690
 KEYWORDS EST.
 SOURCE Snaeda maritima subsp. salsa.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales; Caryophyllales; Chenopodiaceae; Snaeda.
 REFERENCE 1 (bases 1 to 405)
 AUTHORS Zhang, L., Ma, C.L., Wang, P.P., Sun, Y.F., Zhao, Y.X. and Zhang, H.
 TITLE Expressed sequence tags from a halophyte Snaeda salsa cDNA library

JOURNAL

COMMENT

Unpublished (2000)
 Contact: Hui Zhang
 Key Laboratory of Plant Stress Research
 The Biology Department of Shandong Normal University
 No. 88, Wenhua East Road, Jinan, Shandong Province, 250014, PRC
 Tel: (86)531-2960864
 Fax: (86)531-2966954
 Email: zhanghsdnu.edu.cn.

FEATURES

source

Location/Qualifiers

1..405
 /organism="Snaeda maritima subsp. salsa"
 /db_xref="taxon:126914"
 /clone_lib="Snaeda salsa ZAP cDNA library"

/dev_stage="seeding"

/note="Organ: aerial part tissue; Vector: lambda zap;
 Site_1: EcoRI; Site_2: XhoI; total RNA extraction from
 NACL(400mM) treated Snaeda salsa by RNAsent kit(Promega);
 mRNA isolation by MESSAGEMAKER kit(GIBCO BRL);
 directional cDNA synthesis(EcoRI XhoI) by cDNA synthesis
 kit(STRATAGEN); the ZAP express library by GigapackIII
 Gold Cloning kit(STRATAGEN)"

BASE COUNT

112 a 120 c 49 g 123 t 1 others

ORIGIN

Query Match 84.0%; Score 16.8; DB 123; Length 405;
 Best Local Similarity 90.0%; Pred. No. 6; 5e-02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1 gagtaggaagatagaac 20
 ||||| ||||| |||||
 Db 219 GAAGTAGAGGGTAGAACG 200

RESULT 2

LOCUS

DEFINITION

BE859183 411 bp mRNA EST 16-OCT-2000
 SSS0490 Snaeda salsa ZAP cDNA library Snaeda maritima subsp. salsa
 CDNA similar to putative nematode-resistance protein, mRNA
 sequence.

ACCESSION

VERSION

KEYWORDS

ORGANISM

SOURCE

EST

Eukaryota

Viridiplantae

Streptophyta

Embryophyta

Tracheophyta

Spermatophyta

Magnoliophyta

eudicotyledons

core eudicots

Caryophyllales

Caryophyllales

Chenopodiaceae

Snaeda

1 (bases 1 to 411)

Zhang, L., Ma, C.L., Wang, P.P., Sun, Y.F., Zhao, Y.X. and Zhang, H.

TITLE

JOURNAL

COMMENT

Unpublished (2000)

Key Laboratory of Plant Stress Research

The Biology Department of Shandong Normal University

No. 88, Wenhua East Road, Jinan, Shandong Province, 250014, PRC

Tel: (86)531-2960864

Fax: (86)531-2966954

Email: zhanghsdnu.edu.cn.

Location/Qualifiers

1..411

/organism="Snaeda maritima subsp. salsa"

/db_xref="taxon:126914"

/clone_lib="Snaeda salsa ZAP cDNA library"

/dev_stage="seeding"

/note="Organ: aerial part tissue; Vector: lambda zap;
 Site_1: EcoRI; Site_2: XhoI; total RNA extraction from
 NACL(400mM) treated Snaeda salsa by RNAsent kit(Promega);
 mRNA isolation by MESSAGEMAKER kit(GIBCO BRL);
 directional cDNA synthesis(EcoRI XhoI) by cDNA synthesis
 kit(STRATAGEN); the ZAP express library by GigapackIII
 Gold Cloning kit(STRATAGEN)"

BASE COUNT

112 a 123 c 50 g 126 t

REFERENCE	Archosauilia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
AUTHORS	1 (bases 1 to 814) Abdrakhmanov,I., Lodgjin,D., Gerolt,P., Arakawa,H., Law,A., Plachy,J., Korn,B. and Buerstedde,J.M.
TITLE	A large database of chicken buasal ESTs as a resource for the analysis of vertebrate gene function
JOURNAL	Genome Res. 10 (12), 2062-2069 (2000)
MEDLINE	20568495
COMMENT	Contact: Buerstedde JM Cellular Immunology Heinrich-Pette-Institute Martinstr. 52, 20251 Hamburg, Germany Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html. Location/Qualifiers
FEATURES	Source 1..814 /organism="Gallus gallus" /strain="CB" /db_xref="taxon:9031" /cclone="24b16r1" /cclone_1lb="dkfz426" /tissue_type="Bursa of Fabricius" Base Count 227 a 139 c 133 g 315 t Origin
Query Match	84.0%; Score 16.8; DB 104; Length 814;
Best Local Similarity	90.0%; Pred. NO. 7.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
OY	1 gagtaggaagatagaacgc 20
Db	574 GTGTGGAGATGTGAAGG 555
RESULT	6
LOCUS	BE448549/c
DEFINITION	BE448549 196 bp mRNA EST 25-JUL-2000 uc92d01.y1 Soares,mammary_gland_MMLMG Mus musculus cDNA clone
ACCESSION	IMAGE:3369889 5', mRNA sequence.
VERSION	BE448549
KEYWORDS	BE448549.1 GI:9448126
SOURCE	EST. house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)
AUTHORS	Contact: Robert Strausberg, Ph.D. Email: cgaps-re@mail.nih.gov
TITLE	This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. NCI:1079493
JOURNAL	Seq primer: -40RP from Gldco
COMMENT	High quality sequence stop: 159. Location/Qualifiers 1..196 /organism="Mus musculus" /db_xref="taxon:10090" /cclone="IMAGE:3369889" /cclone_1lb="Soares,mammary_gland_MMLMG" /sev="female(lactating)" /tissue_type="mammary gland" /lab_host="DH10B"
FEATURES	/note=Vector: pTT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not

BASE COUNT	40 a	39 c	34 g	83 t
ORIGIN				
Query Match	82.0%;	Score 16.4;	DB 167;	Length 196;
Best Local Similarity	94.4%;	Pred. No. 8.9e+02;		
Matches 17; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
OY	2	agtagaagagatagaac	19	
Db	179	AGTAGAAGAGATAGCAAC	162	
RESULT 7				
AM311691			EST	21-JAN-2000
LOCUS	AM311691	207 bp	mRNA	
DEFINITION	XXNV055F11F Nsf Xylem Normal wood Vertical Pinus taeda cDNA clone			
ACCESSION	XXNV055F11 5', mRNA sequence.			
VERSION	AM311691			
KEYWORDS	AM311691.1 GI:6727337			
SOURCE	EST.			
ORGANISM	loblolly pine.			
REFERENCE	Pinus taeda			
AUTHORS	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
TITLE	Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.			
JOURNAL	1 (bases 1 to 207)			
COMMENT	Sederoff, R.			
	Molecular Basis of Wood Formation in the Pine Megagenome			
	Unpublished (2000)			
	Contact: Johnson, Arthur			
	North Carolina State University			
	Tel: 919 515 7800			
	Fax: 919 515 7801			
	Email: ajohnson@ncsu.edu			
	Seq primer: T3.			
FEATURES				
source	Location/Qualifiers			
	1..207			
	/organism="Pinus taeda"			
	/db_xref="taxon:3352"			
	/clone="XXNV055F11"			
	/clone_1id="Nsf Xylem Normal wood Vertical"			
	/note="Vector: Bluescript SK; Site 1: Eco RI; The			
	sequences contain a 'cDNA adapter' between the EcoRI site			
	and the start of the EST. The adapter sequence is			
	'AATTCGCACGAG' "			
BASE COUNT	68 a	21 c	57 g	58 t 3 others
ORIGIN				
Query Match	82.0%;	Score 16.4;	DB 114;	Length 207;
Best Local Similarity	94.4%;	Pred. No. 9e+02;		
Matches 17; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
OY	1	gagtagaagagatagaac	18	
Db	24	GAGTAGAAGAGATAGAAA	41	
RESULT 8				
BE648384				
LOCUS	BE648384	231 bp	mRNA	EST
DEFINITION	UI-M-BH2.1-aph-e-12-0-UI.r1 NIH BMAP M.S3.1 Mus musculus cDNA clone			
ACCESSION	UI-M-BH2.1-aph-e-12-0-UI 5', mRNA sequence.			
VERSION	BE648384			
KEYWORDS	BE648384.1 GI:9974205			
SOURCE	EST.			
ORGANISM	house mouse.			
	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			


```

/lab host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCACAGACTCTCTTTTCTTTTCTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGAGATTCGAGATTAAATTAATCCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified Bluescript KS(+) after bulk excision from Lambda
FLC I."

```

Query Match	82.0%	Score 16.4	DB 128	Length 252
Best Local Similarity	94.4%	Pred. No. 9.6e+02		
Matches 17, Conservative	0	Mismatches 1	Indels 0	Gaps 0
QY	1	gaatagaagaagataga	aa	18
Db	182	GAGAGGAAGATAGAA		165

RESULT	15
BI155980/c	
LOCUS	293 bp
DEFINITION	musculus cDNA clone A130028422 3', mRNA sequence.
	EST 29-JUN-2000
BI155980	RIKEN full-length enriched, 16 days neonate thymus Mus-

ACCESSION	BB155980
VERSION	BB155980.1
KEYWORDS	GI:8811910
SOURCE	EST.
	house mouse.

ORGANISM Mus musculus

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus
1 (bases 1 to 293)
Konno, H., Alizawa, K., Akahita, S., Akiyama, J., Arakawa, T., Carninci

TITLE
JOURNAL
COMMENT

COMMENT
Contact: Yoshihide Hayashizaki

Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098

Email: genome-reser@rtc.riken.go.jp,
 URL: <http://genome.rtc.riken.go.jp/>
 Carindini, P., Nishiyama, Y., Westover, A., Itoh, M.,
 N. Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 "Thermosensitization and thermoactivation of thermolabile enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)

FEATURES
SOURCE
1. .293
Location/Qualifiers

```

BASE COUNT      65 a      82 c      41 g      104 t      1 others
ORIGIN
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A130028M22"
/clone_lib="RIKEN full-length enriched, 16 days neonate
thymus"
/tissue_type="thymus"
/dev_stage="16 days neonate"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGGAGAGAGAGATCCAGAGGCTCTTTTCTTTTCTTTTCTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGGAGAGAGATTCGAGTTAATTAATTAATTCACCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."
```

Query Match	82.0%	Score 16.4	DB 127	Length 233
Best Local Similarity	94.4%	Pred. No. 9.6e+02		
Matches 17	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	3	qtaagaagatagaacg	20	
db	108	GAAGCAAGCATGAACG	91	

Search completed: October 16, 2001, 18:30:43
Job time: 2476 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 19:00:58 ; Search time 88.04 Seconds
(without alignments)
43.006 Million cell updates/sec

Title: US-09-675-650-4
Perfect score: 20
Sequence: 1 gagtagagagatagaacg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%
Listing First 45 summaries

Database :

Issued_Patents.NA:*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/6C.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	14.8	74.0	67	2	US-08-343-443B-14	Sequence 14, Appl
2	14.8	74.0	750	1	US-08-224-195-1	Sequence 1, Appl
3	14.8	74.0	780	1	US-08-224-195-2	Sequence 2, Appl
4	14.8	74.0	833	2	US-08-343-443B-97	Sequence 97, Appl
5	14.8	74.0	1192	1	US-08-340-539A-3	Sequence 3, Appl
6	14.8	74.0	1192	2	US-08-461-592B-3	Sequence 3, Appl
7	14.8	74.0	1456	3	US-09-308-406-1	Sequence 1, Appl
8	14.8	74.0	1930	1	US-07-982-112-1	Sequence 1, Appl
9	14.8	74.0	2371	2	US-08-343-443B-1	Sequence 1, Appl
10	14.8	74.0	2412	1	US-08-437-027-18	Sequence 18, Appl
11	14.8	74.0	246240	2	US-08-724-394A-20	Sequence 20, Appl
12	14.8	74.0	246240	2	US-08-724-394A-21	Sequence 21, Appl
13	14.8	74.0	246240	2	US-08-724-394A-22	Sequence 22, Appl
14	14.2	71.0	468	4	US-08-821-994-36	Sequence 36, Appl
15	14.2	71.0	507	4	US-08-821-994-75	Sequence 75, Appl
16	14.2	71.0	563	4	US-09-276-531-38	Sequence 38, Appl
17	14.2	71.0	743	4	US-08-821-994-37	Sequence 37, Appl
18	14.2	71.0	743	4	US-08-821-994-38	Sequence 38, Appl
19	14.2	71.0	926	1	US-08-106-507-1	Sequence 1, Appl
20	14.2	71.0	1102	4	US-08-821-994-86	Sequence 86, Appl
21	14.2	71.0	1390	4	US-08-821-994-61	Sequence 61, Appl
22	14.2	71.0	1434	4	US-08-821-994-62	Sequence 62, Appl
23	14.2	71.0	1441	4	US-08-821-994-63	Sequence 63, Appl
24	14.2	71.0	1661	1	US-08-821-994-82	Sequence 82, Appl
25	14.2	71.0	1674	1	US-08-096-762-199	Sequence 199, Appl
26	14.2	71.0	1674	4	US-09-042-353-59	Sequence 59, Appl
27	14.2	71.0	1728	3	US-08-379-802-1	Sequence 1, Appl

28	14.2	71.0	1728	3	US-09-048-129-1	Sequence 1, Appl
29	14.2	71.0	1728	4	US-09-048-079-1	Sequence 1, Appl
30	14.2	71.0	2334	1	US-08-062-632-4	Sequence 4, Appl
31	14.2	71.0	2407	2	US-08-895-522-2	Sequence 2, Appl
32	14.2	71.0	2407	3	US-09-195-391-2	Sequence 2, Appl
33	13.8	69.0	678	5	PCR-US94-04361-24	Sequence 24, Appl
34	13.8	69.0	1121	4	US-09-068-569-1	Sequence 1, Appl
35	13.8	69.0	1718	1	US-07-687-466-1	Sequence 1, Appl
36	13.8	69.0	1718	1	US-08-434-702-1	Sequence 1, Appl
37	13.8	69.0	1718	1	US-08-271-883-1	Sequence 1, Appl
38	13.8	69.0	1718	6	5168064-3	Sequence 1, Appl
39	13.8	69.0	2290	2	US-08-933-821-1	Sequence 1, Appl
40	13.8	69.0	2290	3	US-08-960-507-1	Sequence 1, Appl
41	13.8	69.0	2497	1	US-08-643-219-12	Sequence 12, Appl
42	13.8	69.0	2497	2	US-09-131-995-12	Sequence 12, Appl
43	13.8	69.0	2497	2	US-08-832-087B-12	Sequence 12, Appl
44	13.8	69.0	2497	3	US-08-851-350-12	Sequence 12, Appl
45	13.8	69.0	2497	4	US-09-132-154-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-343-443B-14
Sequence 14, Application US/08343443B
Patent No. 5968734
GENERAL INFORMATION:
APPLICANT: Aurias, Alain
APPLICANT: Delattre, Olivier
APPLICANT: Desmaze, Chantal
APPLICANT: Meiot, Thomas
APPLICANT: Peter, Martine
APPLICANT: Ploougaestel, Beatrice
APPLICANT: Thomas, Gilles
APPLICANT: Zucman, Jessica
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SCAID
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESS:
ADDRESSER: Weiser & Associates
STREET: 230 South Fifteenth Street
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ADIT 1.0 DOS text editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,443B
FILING DATE: 18-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/06123
FILING DATE: 20-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989,6121P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:

LENGTH: 67 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-343-443B-14

Query Match 74.0%; Score 14.8; DB 2; Length 67;
Best Local Similarity 88.9%; Pred. No. 57;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gagtaggaagatagaaa 18
||| ||||| |||||
DB 7 GAGGAGGAAGAGAGAAA 24

RESULT 2
US-08-224-195-1/C
Sequence 1, Application US/08224195
Patent No. 5496705
GENERAL INFORMATION:
APPLICANT: SUGANO, MITSUKO
TITLE OF INVENTION: MONOCLONAL ANTIBODY
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/224.195
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/929.377
FILING DATE: 14-AUG-1992
APPLICATION NUMBER: JP 3-229756
FILING DATE: 16-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5496705man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 39-4161-0S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 750 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Rattus norvegicus
US-08-224-195-1

Query Match 74.0%; Score 14.8; DB 1; Length 750;
Best Local Similarity 88.9%; Pred. No. 78;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 gagtaggaagatagaaa 18
||| ||||| |||||

DB 584 GAGGAGGAAGAGAGAAA 567

RESULT 3
US-08-224-195-2/C
Sequence 2, Application US/08224195
Patent No. 5496705
GENERAL INFORMATION:
APPLICANT: SUGANO, MITSUKO
TITLE OF INVENTION: MONOCLONAL ANTIBODY
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/224.195
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/929.377
FILING DATE: 14-AUG-1992
APPLICATION NUMBER: JP 3-229756
FILING DATE: 16-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5496705man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 39-4161-0S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 780 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: rat
US-08-224-195-2

Query Match 74.0%; Score 14.8; DB 1; Length 780;
Best Local Similarity 88.9%; Pred. No. 79;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gagtaggaagatagaaa 18
||| ||||| |||||
DB 589 GAGGAGGAAGAGAGAAA 572

RESULT 4
US-08-343-443B-97
Sequence 97, Application US/0834343B
Patent No. 5968734
GENERAL INFORMATION:
APPLICANT: Aurias, Alain
APPLICANT: Delattre, Olivier
APPLICANT: Desmaze, Chantal
APPLICANT: Melot, Thomas
APPLICANT: Peter, Martine
APPLICANT: Ploougasstel, Beatrice

APPLICANT: Thomas, Gilles
APPLICANT: Zucman, Jessica
TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
TITLE OF INVENTION: TRANSLATIONS
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weiser & Associates
STREET: 230 South Fifteenth Street
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: AEDIT 1.0 DOS text editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,443B
FILING DATE: 18-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/06123
FILING DATE: 20-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-343-443B-97

Query Match 74.0%; Score 14.8; DB 2; Length 833;
Best Local Similarity 88.9%; Pred. No. 79;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gagtaggaagatagaa 18
||| ||||| |||||
DB 575 GAGGAGGAGAGAGAAA 592

RESULT 5
US-08-340-539A-3
Sequence 3, Application US/08340539A
Patent No. 5808025
GENERAL INFORMATION:
APPLICANT: Tedder, Thomas F.
APPLICANT: Kansas, Geoffrey S.
TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS
TITLE OF INVENTION: BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,539A
FILING DATE: 16-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/008,459
FILING DATE: 25-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gunnison, Jane
REGISTRATION NUMBER: 38,479
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1192 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-340-539A-3

Query Match 74.0%; Score 14.8; DB 1; Length 1192;
Best Local Similarity 88.9%; Pred. No. 83;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 gtgagaagatagaaacg 20
||| ||||| |||||
DB 700 GAAGGAGGATGAGAAAG 717

RESULT 6
US-08-461-592B-3
Sequence 3, Application US/08461592B
Patent No. 5834425
GENERAL INFORMATION:
APPLICANT: Tedder, Thomas F.
APPLICANT: Kansas, Geoffrey S.
TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS
TITLE OF INVENTION: BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,592B
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/340,539
FILING DATE: 16-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/008,459
FILING DATE: 25-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: James F. Haley, Jr.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: CG-104

```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 596-9000
: TELEFAX: (212) 596-9090
: TELEX: 14-8367
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1192 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: US-08-461-592B-3

Query Match      74.0%: Score 14.8; DB 2; Length 1192;
Best Local Similarity 88.9%: Pred. No. 83;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      3 gtaggaagatagaacg 20
      | ||||| ||||| |||||
Db      700 GAGGAGAGATGAAAG 717

RESULT      7
US-09-308-406-1
: Sequence 1, Application US/09308406
: Patent No. 6159666
: GENERAL INFORMATION:
: APPLICANT: Dijkema, Rein
: APPLICANT: van den Wijngaard, Arthur
: TITLE OF INVENTION: Method and materials for the screening of therapeutic
: TITLE OF INVENTION: agents for the prevention and/or treatment of
: TITLE OF INVENTION: osteoporosis
: FILE REFERENCE: I/96227 US
: CURRENT APPLICATION NUMBER: US/09/308,406
: CURRENT FILING DATE: 1999-06-21
: EARLIER APPLICATION NUMBER: PCT/EP97/06668
: EARLIER FILING DATE: 1997-11-20
: EARLIER APPLICATION NUMBER: EP96203283.5
: EARLIER FILING DATE: 1996-11-22
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 1456
: TYPE: DNA
: ORGANISM: human
: US-09-308-406-1

Query Match      74.0%: Score 14.8; DB 3; Length 1456;
Best Local Similarity 88.9%: Pred. No. 86;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 gagtagaagatagaac 18
      ||| ||||| ||||| |||||
Db      1122 gagagagaagaagaa 1139

RESULT      8
US-07-982-112-1/c
: Sequence 1, Application US/07982112
: Patent No. 5346813
: GENERAL INFORMATION:
: APPLICANT: BODENMULLER, Heinz
: APPLICANT: DESSAUER, Andreas
: TITLE OF INVENTION: METHOD FOR THE DETECTION OF A SMALL
: TITLE OF INVENTION: CELL LUNG CARCINOMA
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nikaido, Marmelstein, Murray & Oram
: STREET: 655 Fifteenth Street N.W. Suite 330
```

```

: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20005-5701
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/982,112
: FILING DATE: 19921125
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/617,102
: FILING DATE: 21-NOV-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Chin, Monica F.
: REGISTRATION NUMBER: 36,105
: REFERENCE/DOCKET NUMBER: 564-1906
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)638-5000
: TELEFAX: (202)638-4810
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1930 base pairs
:   TYPE: NUCLEIC ACID
:   STRANDEDNESS: double
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 10..933
: US-07-982-112-1

Query Match      74.0%: Score 14.8; DB 1; Length 1930;
Best Local Similarity 88.9%: Pred. No. 89;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 gagtagaagatagaac 18
      ||| ||||| ||||| |||||
Db      1309 GAGGAGAGAGGAGGAAA 1292

RESULT      9
US-08-343-443B-1
: Sequence 1, Application US/0834343B
: Patent No. 5968734
: GENERAL INFORMATION:
: APPLICANT: Aurias, Alain
: APPLICANT: Delattre, Olivier
: APPLICANT: Desmaze, Chantal
: APPLICANT: Melot, Thomas
: APPLICANT: Peter, Martine
: APPLICANT: Ploougastel, Beatrice
: APPLICANT: Thomas, Gilles
: APPLICANT: Zucman, Jessica
: TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
: TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
: TITLE OF INVENTION: TRANSLOCATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
: TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
: NUMBER OF SEQUENCES: 129
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Welser & Associates
: STREET: 230 South Fifteenth Street
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: AEDIT 1.0 DOS text editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,443B
FILING DATE: 18-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/06123
FILING DATE: 20-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989.6121P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2371 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 25..1992
US-08-343-443B-1

Query Match 74.0%; Score 14.8; DB 2; Length 2371;
Best Local Similarity 88.9%; Pred. No. 91;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gagtaggaagatagaa 18
||| ||||| |||||
Db 7 GAGGAGGAGGAGAGAA 24

RESULT 10
US-08-437-027-18
Sequence 18, Application US/08437027
Patent No. 5670317
GENERAL INFORMATION:
APPLICANT: Landanyi, Marc
TITLE OF INVENTION: A DIAGNOSTIC TEST FOR TEST FOR THE DESMOPLASTIC
TITLE OF INVENTION: SMALL ROUND CELL TUMOR
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/437,027
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 46416/JPM/CCA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400

TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2412 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-437-027-18

Query Match 74.0%; Score 14.8; DB 1; Length 2412;
Best Local Similarity 88.9%; Pred. No. 91;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gagtaggaagatagaa 18
||| ||||| |||||
Db 7 GAGGAGGAGGAGAGAA 24

RESULT 11
US-08-724-394A-20
Sequence 20, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Flits, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note="HLA-H.COMTIG"
US-08-724-394A-20

Query Match 74.0%; Score 14.8; DB 2; Length 246240;

Best Local Similarity 88.9%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 2 agtaggaagatagaac 19
|||||
Db 141872 AGTAGGAAGATTAATAC 141889

RESULT 12
US-08-724-394A-21
Sequence 21, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-21

Query Match 74.0%; Score 14.8; DB 2; Length 246240;
Best Local Similarity 88.9%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 2 agtaggaagatagaac 19
|||||
Db 141872 AGTAGGAAGATTAATAC 141889
RESULT 13
US-08-724-394A-22
Sequence 22, Application US/08724394A
Patent No. 5872237

GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-22

Query Match 74.0%; Score 14.8; DB 2; Length 246240;
Best Local Similarity 88.9%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 2 agtaggaagatagaac 19
|||||
Db 141872 AGTAGGAAGATTAATAC 141889
RESULT 14
US-08-821-994-36
Sequence 36, Application US/08821994A
Patent No. 6228643
GENERAL INFORMATION:
APPLICANT: Greenland, Andrew J
APPLICANT: Thomas, Didier RP
APPLICANT: Jepson, Ian
TITLE OF INVENTION: Promoters
FILE REFERENCE: PPD 50108
CURRENT APPLICATION NUMBER: US/08/821,994A
EARLIER FILING DATE: 1997-03-22
EARLIER APPLICATION NUMBER: PCT/GB97/00729
EARLIER FILING DATE: 1997-03-18
EARLIER APPLICATION NUMBER: GB 9606062.9
EARLIER FILING DATE: 1996-03-22

```

; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Brassica napus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(29), 439, 441
; OTHER INFORMATION: n is unknown
US-08-821-994-36

```

```

Query Match      71.0%; Score 14.2; DB 4; Length 468;
Best Local Similarity 84.2%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 agtaggaagatagaacg 20
      1 ||||| ||| |||||
Db      391 attagatcgagagaacg 409

```

```

RESULT 15
US-08-821-994-75
; Sequence 75, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 75
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(5), 10, 13, 15, 16, 28, 150, 276, 285, 292, 312, 337
; OTHER INFORMATION: n is uncertain
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 355, 357, 447, 477, 495
; OTHER INFORMATION: n is uncertain
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Consensus
US-08-821-994-75

```

```

Query Match      71.0%; Score 14.2; DB 4; Length 507;
Best Local Similarity 84.2%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 agtaggaagatagaacg 20
      1 ||||| ||| |||||
Db      394 attagatcgagagaacg 412

```

Search completed: October 16, 2001, 19:01:18
Job time: 4096 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 19:01:36 ; Search time 164.45 Seconds
(without alignments)
76.364 Million cell updates/sec

Title: US-09-675-650-4

Perfect score: 20
Sequence: 1 gagtagaagatagaacac 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq.0601.*
1: /SIDSB/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSB/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSB/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSB/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSB/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSB/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSB/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSB/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSB/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSB/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSB/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSB/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSB/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSB/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSB/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSB/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSB/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSB/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSB/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSB/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSB/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSB/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	16.8	84.0	792	20	AAZ12053
2	16.8	84.0	6147	21	AAZ53858
3	16.4	82.0	347	21	AAZ19786
4	15.8	79.0	300	20	AAZ13970
5	15.8	79.0	1926	21	AAZ08075
6	15.8	79.0	2720	22	AAZ02348
7	15.8	79.0	2844	21	AAZ98219
8	15.8	79.0	611590	21	AAZ22303
9	15.4	77.0	6761	20	AAZ20517
10	15.2	76.0	485	20	AAV89036
11	15.2	76.0	620	21	AAZ09941

12	15.2	76.0	1544	18	AAZ86839
13	15.2	76.0	3127	20	AAZ40171
14	15.2	76.0	6774	21	AAZ39174
15	15.2	76.0	144460	21	AAZ93815
16	15.2	76.0	1038602	20	AAZ01425
17	15.2	76.0	300	20	AAZ13594
18	15.2	76.0	300	20	AAZ13588
19	14.8	74.0	233	22	AAZ44494
20	14.8	74.0	393	21	AAZ07987
21	14.8	74.0	401	21	AAZ28961
22	14.8	74.0	454	21	AAZ01659
23	14.8	74.0	735	4	AAZ30161
24	14.8	74.0	750	14	AAZ03214
25	14.8	74.0	760	22	AAZ22877
26	14.8	74.0	763	4	AAZ30162
27	14.8	74.0	763	11	AAZ006498
28	14.8	74.0	763	19	AAZ33298
29	14.8	74.0	814	21	AAZ75898
30	14.8	74.0	833	14	AAZ050688
31	14.8	74.0	876	4	AAZ30108
32	14.8	74.0	876	11	AAZ006497
33	14.8	74.0	876	19	AAZ33297
34	14.8	74.0	977	6	AAZ50357
35	14.8	74.0	1192	15	AAZ071007
36	14.8	74.0	1342	21	AAZ49714
37	14.8	74.0	1346	21	AAZ41864
38	14.8	74.0	1443	19	AAZ94208
39	14.8	74.0	1456	19	AAZ32783
40	14.8	74.0	1533	21	AAZ49698
41	14.8	74.0	1535	21	AAZ40036
42	14.8	74.0	1686	2	AAZ10014
43	14.8	74.0	2371	14	AAZ050643
44	14.8	74.0	2403	18	AAZ97870
45	14.8	74.0	2412	18	AAZ97869

ALIGNMENTS

RESULT 1	AAZ12053/C	standard; DNA; 792 BP.
ID	AAZ12053	
AC	AAZ12053	
XX		
DT	08-OCT-1999	(first entry)
XX		
DE	Neisseria gonorrhoeae complete ORF31 sequence.	
XX		
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;	
KW	treatment; Neisseria infection; meningitis; septicaemia; gonorrhea; ss.	
XX		
OS	Neisseria gonorrhoeae.	
XX		
PN	W09924578-A2.	
XX		
PD	20-MAY-1999.	
XX		
PF	09-OCT-1998;	98MO-IB01665.
XX		
PR	01-SEP-1998;	98GB-0019016.
PR	06-NOV-1997;	97GB-0023516.
PR	14-NOV-1997;	97GB-0024190.
PR	18-NOV-1997;	97GB-0024386.
PR	27-NOV-1997;	97GB-0025158.
PR	10-DEC-1997;	97GB-0026147.
PR	14-JAN-1998;	98GB-0000759.
XX		
PA	(CHIR-) CHIRON SPA.	
XX		
PI	Grandi G, Masiagnanl V, Pizza M, Rappuoli R, Scarlato V;	
XX		
DR	WPI, 1999-327407/27.	

cDNA encoding rat
W09904265 Seq ID N
Bone morphogenetic
Olfactory receptor
Complete genome se
Human gene express
Human gene express
Mouse dextran sodi
Human secreted pro
Human secreted pro
Human secreted pro
Sequence encoding
Rat synaptophysin
Human prostate can
Sequence of the B.
IFN-alpha 76 gene
Escherichia coli t
Human ORF1453
Human Ews gene pro
Sequence encoding
IFN-alpha 76 gene.
DNA encoding Inter
Human Interferon-a
Ilym-1 gene exon I
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Human BMP-4 promot
Arabidopsis thalia
Arabidopsis thalia
Sequence encoding
Human Ews gene clo
cDNA encoding a ch
cDNA encoding a ch

DR P-PSDB: AAY38591.
 XX
 PT Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for
 PT diagnosis, treatment and prevention of infection
 XX
 PS Claim 9; Page 150-151; 524pp; English.
 CC Nucleotide sequences AA21972-212358 represent open reading frames
 CC (ORFs) of *Neisseria meningitidis* and *N. gonorrhoeae* which encode
 CC antigenic proteins (see AAY38499-Y38944). The antigenic proteins, their
 CC fragments, their nucleic acids and antibodies are used for diagnosis,
 CC prevention (as vaccines) or treatment of *Neisseria* infections,
 CC such as meningitis, septicemia and gonorrhea. Both organisms
 CC are closely related. Fragments of the nucleic acids are useful
 CC as hybridisation probes and antisense reagents.
 CC
 XX
 SQ Sequence 792 BP; 210 A; 184 C; 189 G; 209 T; 0 other;

Query Match 84.0%; Score 16.8; DB 20; Length 792;
 Best Local Similarity 90.0%; Pred. No. 57;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gagtagaagatagaacg 20
 |||||
 DB 148 GAGTAGGAATGAAAGAACG 129

RESULT 2
 AA23858/c
 ID AA23858 standard; DNA; 6147 BP.
 XX
 AC AA23858;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE *Neisseria gonorrhoeae* ORF 564 partial DNA sequence SEQ ID NO:1665.
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy; ds.
 XX
 OS *Neisseria gonorrhoeae*.
 XX
 PN MO957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PE 30-APR-1999; 99WO-US09346.
 XX
 PF 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR P-PSDB: AAY75096.
 XX
 PT Novel *Neisseria* polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX
 PS Claim 7; Page 849-851; 1453pp; English.

XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides: AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 CC
 XX
 SQ Sequence 6147 BP; 1998 A; 1476 C; 1444 G; 1229 T; 0 other;

Query Match 84.0%; Score 16.8; DB 21; Length 6147;
 Best Local Similarity 90.0%; Pred. No. 62;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gagtagaagatagaacg 20
 |||||
 DB 148 GAGTAGGAATGAAAGAACG 129

RESULT 3
 AAC19786
 ID AAC19786 standard; cDNA; 347 BP.
 XX
 AC AAC19786;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein 5' EST, SEQ ID NO: 23861.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 XX
 OS *Homo sapiens*.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PE 21-FEB-2000; 2000EP-0200610.
 XX
 PF 26-FEB-1999; 99US-0122487.
 XX
 PG (GEST) GENSET.
 XX
 PA Dunas Mline Edwards J, Duclert A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures
 XX
 PS Claim 1; SEQ ID 23861; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from
 CC cDNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.

XX Sequence 347 BP; 132 A; 51 C; 57 G; 105 T; 2 other;

Query Match 82.0%; Score 16.4; DB 21; Length 347;

Best Local Similarity 94.4%; Pred. No. 83;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 agtaggaagatagaac 19
|||||

DB 246 agtaggaagacagaac 263

RESULT 4
AA213970
ID AA213970 standard; cDNA; 300 BP.

XX AA213970;

DE 12-OCT-1999 (first entry)

Human gene expression product cDNA sequence SEQ ID NO:1439.

Human; gene; gene expression product; diagnosis; therapy; probe;
detection; mapping; tissue typing; profiling; forensic; cancer;
genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

PN W0938972-A2.

OS 05-AUG-1999.

PF 28-JAN-1999; 99WO-US01619.

PR 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

XX (CHTR) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.

XX Sequence 300 BP; 105 A; 42 C; 58 G; 95 T; 0 other;

Query Match 79.0%; Score 15.8; DB 20; Length 300;

Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 agtaggaagatagaac 20
|||||

DB 16 aataggagagatagaac 34

RESULT 5
AAA08075/c

ID AAA08075 standard; cDNA to mRNA; 1926 BP.

XX AAA08075;

AC 21-JUN-2000 (first entry)

Human NPw38BP1 transcription related protein encoding cDNA SEQ ID NO:2.

Human; NPw38BP1; NPw38; transcription related protein; diagnosis;

genetic disease; gene therapy; antitumor; ds.

OS Homo sapiens.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

The present invention describes a library of human polynucleotides
comprising the sequences given in AA212532 to AA217779. Also described is
a method of detecting differentially expressed genes correlated with the
cancerous state of a mammalian cell, comprising detecting at least one
differentially expressed gene product in a test sample from a cell
suspected of being cancerous, where the gene product is encoded by one
of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
polynucleotides can be used as a source of primers and probes, which can
be used for a variety of purpose, e.g. detection of expression levels,
mapping, tissue typing or profiling, forensics, genetic analysis and
detection of polymorphisms. Polypeptides encoded by the polynucleotides
can be used for raising antibodies for experimental, diagnostic and
therapeutic purposes. The polynucleotides may also be used to construct

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 79.0%; Score 15.8; DB 21; Length 1926;

```
Oy      1 gagtaggaagatagaac 19
        || |||||
Db      556 GAAGAGGAGGATGAGAAC 538

RESULT  6
ID      AAD02348/c
        AAD02348 standard; cDNA; 2720 BP.
XX
AC      AAD02348;
XX
DT      24-APR-2001 (first entry)
XX
DE      Human RNA metabolism protein (RMEP-5) cDNA.

KW      Human; RNA metabolism protein; RMEP; noctropic; neuroleptic; antileuc;
KW      tranquilliser; antianaemic; antidiabetic; immunosuppressive; cytostatic;
KW      antisthmatic; antiinflammatory; anti-HIV; human immunodeficiency virus;
KW      antiarthritic; antiatherosclerotic; antiatherosclerotic; anti allergic;
KW      antirheumatoid; antiparkinsonian; antihypoid; nephrotrophic; antigout;
KW      thyromimetic; RMEP expression modulator; transgenic; spinal cord disease;
KW      nervous system disorder; Alzheimer's disease; therapy; gene therapy;
KW      neuromuscular disorder; hepatitis; cancer; cell proliferative disorder;
KW      peripheral nervous system disorder; cirrhosis; cranial nerve disorder;
KW      progressive neural autonomic nervous system disorder; Addison's disease;
KW      amyotrophic lateral sclerosis; autoimmune disorder; drug screening; ss.
XX
OS      Homo sapiens.
XX
FH      Key
FT      192..2117
FT      CDS
FT      /tag= a
FT      /product= "Human RNA metabolism protein (RMEP-5)"

XX      WO200078952-A2.
XX
XX      28-DEC-2000.
XX
PD      15-JUN-2000; 2000WO-US16644.
XX
XX      17-JUN-1999; 99US-0139922.
XX
XX      (INCY-) INCYTE GENOMICS INC.
XX      (BAUG/) BAUGHN M R.
XX
PI      Bandman O, Yue H, Lal P, Tang YT, Reddy R, Azimzal Y;
XX
XX      WPI: 2001-102723/11.
XX      P-PSDB: NAY72165.
XX
XX      New human RNA metabolism proteins (RMEP), useful for diagnosing,
XX      treating, preventing nervous system, cell proliferative,
XX      autoimmune/inflammatory disorders associated with abnormal expression
XX      of RMEP.
XX
XX      Example 1; Page 97-98; 103pp; English.
XX
XX      The present sequence is human RNA metabolism protein (RMEP-5) cDNA
XX      (Clone ID 1858421) obtained from PROSNOR18 cDNA library.
XX      Agonists and antagonists of RMEP cDNA are useful for treating diseases or
XX      conditions associated with altered expression of functional RMEP. RMEP
XX      sequence or their mammalian homologues are useful for creating 'knock
XX      out' or 'knock in' humanised animals or transgenic animals to model
XX      human disease. RMEP sequence is useful in the diagnosis, prevention and
XX      treatment of nervous system disorders e.g. Alzheimer's disease, Pick's
XX      disease, Huntington's disease, Parkinson's disease, amyotrophic lateral
XX      sclerosis, and other motor neuron disorders, progressive neural autonomic
XX      nervous system disorders, cranial nerve disorders, spinal cord diseases,
XX      muscular dystrophy and other neuromuscular disorders, peripheral nervous
XX      system disorders, mental disorders including anxiety and schizophrenia,
XX      amnesia etc, cell proliferative disorders e.g. actinic keratosis,
XX      arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, mixed connective
```

```
CC      tissue disease (MCTD), cancers e.g. adenocarcinoma, leukemia, lymphoma,
CC      melanoma etc., and autoimmune/inflammatory disorders such as acquired
CC      immuno deficiency syndrome (AIDS), Addison's disease, allergies, anaemia,
CC      asthma, diabetes mellitus, rheumatoid arthritis, Grave's disease and
CC      autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED).
CC      RMEP cDNA is useful for somatic or germ-line gene therapy. RMEP sequence
CC      is useful several drug screening assays.
XX
XX      Sequence 2720 BP; 781 A; 669 C; 604 G; 666 T; 0 other;
XX
XX      Query Match
XX      Best Local Similarity 79.0%; Score 15.8; DB 22; Length 2720;
XX      Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
Oy      1 gagtaggaagatagaac 19
        || |||||
Db      747 GAAGAGGAGGATGAGAAC 729

RESULT  7
ID      AAC98219/c
        AAC98219 standard; cDNA; 2844 BP.
XX
AC      AAC98219;
XX
DT      09-MAR-2001 (first entry)
XX
DE      Human colon cancer antigen nucleotide sequence SEQ ID NO:229.
XX
XX      Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX      identification; cytostatic; cardioactive; neuroprotective; vulnerrary;
XX      immunomodulatory; muscular; gynaecological; gastrointestinal;
XX      nephrotrophic; antiinfective; antibacterial; gene therapy; wound;
XX      neural disorder; immune system disorder; muscular disorder;
XX      reproductive disorder; gastrointestinal disorder; renal disorder;
XX      infectious disease; cardiovascular disorder; ss.
XX
XX      Homo sapiens.
XX
OS      WO200055351-A1.
XX
XX      21-SEP-2000.
XX
PD      08-MAR-2000; 2000WO-US05883.
XX
XX      12-MAR-1999; 99US-0124270.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX
XX      Rosen CA, Ruben SM;
XX
XX      WPI: 2000-587534/55.
XX      P-PSDB: AAB53462.
XX
XX      Colon cancer associated gene sequences, referred to as colon cancer
XX      antigens, useful for the treatment, prevention, and diagnosis of colon
XX      disorders such as colon cancer.
XX
XX      Claim 1; Page 650-651; 2104pp; English.
XX
XX      AAC97991 to AAC98763 encode the human colon cancer associated proteins,
XX      called human colon cancer antigens, given in AAB53234 to AAB54006. The
XX      human colon cancer antigens can have cytostatic, cardioactive, muscular;
XX      neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
XX      vulnerrary, nephrotrophic, antiinfective and antibacterial activities, and
XX      can be used in gene therapy. The colon cancer antigen polynucleotides,
XX      proteins and antibodies to the proteins are useful for the prevention,
XX      treatment and diagnosis of colon disorders, such as colon cancer. The
XX      polynucleotides may be used in diagnostics and research, such as for
XX      chromosome identification, and as hybridisation probes. The proteins
XX      may also be used to prevent diseases such as neural disorders, immune
XX      system disorders, muscular disorders, reproductive disorders,
```

CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.

XX Sequence 2844 BP; 816 A; 691 C; 625 G; 711 T; 1 other;

Query Match 79.0%; Score 15.8; DB 21; Length 2844;
Best Local Similarity 89.5%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gagtaggaagatagaac 19
||| ||||| ||||| |||||
Db 743 GAGAGGAAAGATAGAAC 725

RESULT 8

AAF22303/C
ID AAF22303 standard; DNA; 611590 BP.

XX AAF22303;

XX 20-MAR-2001 (first entry)

XX Arabidopsis thaliana chromosome 2 centromere.

XX Centromere; michrosome; vector; ds.

XX Arabidopsis thaliana.

XX WO200055325-A2.

XX 21-SEP-2000.

XX 17-MAR-2000; 2000WO-US07392.

XX 18-MAR-1999; 99US-0125219.

XX 01-APR-1999; 99US-0127409.

XX 18-MAY-1999; 99US-0134770.

XX 13-SEP-1999; 99US-0153584.

XX 17-SEP-1999; 99US-0154603.

XX (UYCH-) UNIV CHICAGO.

XX Preuss D, Copenhaver G, Keith K;

XX WPI; 2000-587529/55.

XX Reclombant DNA construct comprising a plant centromere, useful for
XX the construction of transgenic plant and animal cells

XX Claim 45; Page 820-959; 1449pp; English.

CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited michrosomes which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.

XX Sequence 611590 BP; 181893 A; 124460 C; 120254 G; 184983 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 611590;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gagtaggaagatagaac 19
||| ||||| ||||| |||||
Db 476561 GAGAGGAAAGATAGAAC 476543

RESULT 9

AA20517
ID AAX20517 standard; DNA; 6761 BP.

XX AAX20517;

XX 05-MAY-1999 (first entry)

XX Polynucleotide sequence from the genome of Treponema pallidum.

XX Treponema pallidum infection; syphilis; Borrelia infection; animal;
XX enzyme production; ds.

XX Treponema pallidum.

XX WO9859034-A2.

XX 30-DEC-1998.

XX 23-JUN-1998; 98WO-US13041.

XX 24-JUN-1997; 97US-0050667.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Fraser CM;

XX WPI; 1999-081273/07.

XX New isolated Treponema pallidum nucleic acids - used to develop
XX products for the detection, diagnosis, characterization, prevention
XX and therapy of T. pallidum infections, particularly syphilis

XX Claim 1; Page 285-289; 1150pp; English.

CC AAX20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterization, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.

XX Sequence 6761 BP; 1693 A; 1585 C; 1749 G; 1723 T; 11 other;

Query Match 77.0%; Score 15.4; DB 20; Length 6761;
Best Local Similarity 94.1%; Pred. No. 2.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gagtaggaagatagaac 17
||| ||||| ||||| |||||
Db 292 gagtaggaagatagaac 308

XX 12-FEB-1999 (first entry)

XX EST clone IQ55.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;
XX tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
XX chemotaxis; chemokines; haemostasis; gene therapy; thrombolysis;
XX receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX Homo sapiens.

XX WO9845437-A2.

PD	XX	15-OCt-1998.
PF	XX	10-APR-1998; 98WO-US06956.
PI	XX	10-APR-1997; 97US-0837312.
PS	XX	(GEMV) GENETICS INST INC.
PT	XX	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PY	XX	Racie LA, Spaulding V, Treacy M,
RA	XX	WPI; 1999-070078/06.
RB	XX	New polynucleotides encoding human secreted proteins - derived from
RC	XX	e.g. human blood, kidney, foetal lung, placenta, testes, brain,
RD	XX	ovary, pituitary, retina and colon CDNA libraries
RE	XX	Claim 1; Page 595-596; 641pp; English.
RF	XX	The present sequence represents an expressed sequence tag (EST), and is
RG	XX	a polynucleotide of the invention. The polynucleotides of the invention
RH	XX	are all secreted EST sequences isolated from a variety of human tissue
RI	XX	sources. The EST sequences and proteins encoded by them are predicted to
RJ	XX	have useful biological activities which would make them suitable for
RK	XX	treating, preventing or ameliorating medical conditions in humans and
RL	XX	animals, although no supporting data is given. Suggested activities
RM	XX	include nutritional activity, immune stimulating or suppressing activity,
RN	XX	hematopoiesis regulating activity, tissue growth activity,
RO	XX	activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic
RP	XX	and thrombolytic activity, receptor/ligand activity, anti-inflammatory
RQ	XX	activity, cachectin/tumour invasion suppressor activity, tumour inhibition
RS	XX	therapy. The EST sequences are also stated to be useful for gene
RT	XX	
RU	XX	Sequence 485 BP; 204 A; 58 C; 117 G; 106 T; 0 other;
RV	XX	
RW	XX	
RX	XX	
RY	XX	
RZ	XX	
SA	XX	
SB	XX	
SC	XX	
SD	XX	
SE	XX	
SF	XX	
SG	XX	
SH	XX	
SI	XX	
SJ	XX	
SK	XX	
SL	XX	
SM	XX	
SN	XX	
SO	XX	
SP	XX	
SQ	XX	
SR	XX	
SS	XX	
ST	XX	
SV	XX	
SW	XX	
SX	XX	
SY	XX	
SZ	XX	
TA	XX	
TB	XX	
TC	XX	
TD	XX	
TE	XX	
TF	XX	
TF	XX	
TH	XX	
TI	XX	
TJ	XX	
TK	XX	
TL	XX	
TM	XX	
TN	XX	
TO	XX	
TP	XX	
TP	XX	
TQ	XX	
TR	XX	
TS	XX	
TT	XX	
TU	XX	
TU	XX	
TV	XX	
TV	XX	
TX	XX	
TY	XX	
TZ	XX	
UA	XX	
UB	XX	
UC	XX	
UD	XX	
UE	XX	
UF	XX	
UG	XX	
UH	XX	
UI	XX	
UI	XX	
IJ	XX	
UK	XX	
UL	XX	
UM	XX	
UN	XX	
UN	XX	
UP	XX	
UP	XX	
UQ	XX	
UR	XX	
US	XX	
UT	XX	
UU	XX	
UV	XX	
UV	XX	
VX	XX	
VY	XX	
VZ	XX	
WA	XX	
WB	XX	
WC	XX	
WD	XX	
WE	XX	
WF	XX	
WG	XX	
WH	XX	
WI	XX	
WI	XX	
XJ	XX	
XK	XX	
XL	XX	
XM	XX	
XN	XX	
XO	XX	
XP	XX	
XP	XX	
XQ	XX	
XR	XX	
XS	XX	
XT	XX	
YT	XX	
YU	XX	
YV	XX	
YW	XX	
YZ	XX	
ZA	XX	
ZB	XX	
ZC	XX	
ZD	XX	
ZE	XX	
ZF	XX	
ZG	XX	
ZH	XX	
ZI	XX	
ZJ	XX	
ZK	XX	
ZL	XX	
ZM	XX	
ZN	XX	
ZO	XX	
ZP	XX	
ZP	XX	
ZQ	XX	
ZR	XX	
ZS	XX	
ZT	XX	
ZU	XX	
ZV	XX	
ZV	XX	
ZZ	XX	
ZA	XX	
ZB	XX	
ZC	XX	
ZD	XX	
ZE	XX	
ZF	XX	
ZG	XX	
ZH	XX	
ZI	XX	
ZJ	XX	
ZK	XX	
ZL	XX	
ZM	XX	
ZN	XX	
ZO	XX	
ZP	XX	
ZP	XX	
ZQ	XX	
ZR	XX	
ZS	XX	
ZT</		

PA	(NOVO) NOVO NORDISK BIOTECH INC.
PA	(NOVO) NOVO NORDISK AS.
PI	Berka RM, Rey MW, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
PI	WPL; 2000-594572/56.
DR	
XX	
PT	Monitoring differential expression of genes in filamentous fungal cells
PT	uses fluorescence-labeled nucleic acids isolated from the cells and a
PT	substrate of expressed sequence tags -
PS	Claim 86; Page 1287; 3161pp; English.
XX	
CC	The present invention describes a method for monitoring differential
CC	expression of genes in a first filamentous fungal (FF) cell relative to
CC	expression of the same genes in one or more second filamentous fungal
CC	cells. The method uses fluorescence-labeled nucleic acids isolated from
CC	the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC	are used in the methods for monitoring differential expression of genes
CC	in a first filamentous fungal (FF) cell relative to expression of the
CC	same genes in one or more second filamentous fungal cells. Monitoring
CC	the global expression of genes from FF cells allows the production
CC	potential of the microorganisms to be improved. New genes may be
CC	discovered, possible functions of unknown open reading frames can be
CC	identified and gene copy number variation and stability can be
CC	monitored. The expression of genes can be used to study how FF cells
CC	adapt to changes in culture conditions, environmental stress, spore
CC	morphogenesis, recombination, metabolic or catabolic pathway
CC	engineering. Using ESTs provides several advantages over genomic or
CC	random cDNA clones including elimination of redundancy as one spot on an
CC	array equals one gene or open reading frame, and organisation of the
CC	microarrays based on function of the gene products to facilitate
CC	analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC	Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC	niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC	AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC	all specifically claimed in the present invention.
XX	
SQ	Sequence 620 BP; 173 A; 110 C; 143 G; 189 T; 5 other:
OY	Query Match 76.0%; Score 15.2; DB 21; Length 620; Best Local Similarity 85.0%; Pred. No. 3e+02; Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
DB	1 gagtagaagaagatagaacg 20 177 gagaagaagaagtgcgaacg 196
RESULT 12	
AAT86839/C	
ID	AAT86839 standard; CDNA; 1544 BP.
AC	AAT86839;
DT	19-DEC-1997 (first entry)
DE	CDNA encoding rat CC chemokine receptor.
KX	rat; CC chemokine receptor; screen; binding; ligand; ds.
OS	Rattus rattus.
XX	
FH	Key Location/Qualifiers
FT	CDS 33..1100
FT	/tag= a
PN	JP09227599-A.
PD	02-SEP-1997.
PF	22-FEB-1996; 96JUP-0035192.

XX 22-FEB-1996; 96JP-0035192.
XX (TAKE) TAKEDA CHEM IND LTD.
XX WPI; 1997-486426/45.
XX P-PSDB; AAM29179.
XX CC Chemokine receptor protein - useful to screen for novel binding
XX compounds
XX Claim 4; Page 21-22; 26pp; Japanese.
XX This cDNA encodes a rat CC chemokine receptor. The receptor can be used
XX to screen for novel binding compounds and for preparation of antibodies
XX or antiserum.
XX Sequence 1544 BP; 415 A; 347 C; 330 G; 452 T; 0 other;
XX
XX Query Match 76.0%; Score 15.2; DB 18; Length 1544;
XX Best Local Similarity 85.0%; Pred. No. 3.1e+02;
XX Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 gagtaggaagatagaacg 20
DB 785 GAGGAGGAAGATAGACCG 766
||| ||||| ||||| ||
RESULT 13
AAZ40171
ID AAZ40171 standard; DNA; 3127 BP.
XX AAZ40171;
XX 02-JUL-1999 (first entry)
XX WO9904265 Seq ID No: 683.
XX Cancer associated antigen; diagnosis; research; treatment; human;
XX breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
XX prostate cancer; ss.
XX Homo sapiens.
XX WO9904265-A2.
XX 28-JAN-1999.
XX 15-JUL-1998; 98WO-US14679.
XX 22-JUN-1998; 98US-0102322.
XX 17-JUL-1997; 97US-0896164.
XX 10-OCT-1997; 97US-0061599.
XX 10-OCT-1997; 97US-0061765.
XX 10-OCT-1997; 97US-0948705.
XX 11-OCT-1997; 97GB-0021697.
XX (LUDWIG) INST CANCER RES.
XX Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ;
XX Pfeundschnuh M, Sahin U, Scanlan MJ, Stockert E;
XX Tureci O;
XX WPI; 1999-132448/11.
XX New isolated cancer associated nucleic acids and polypeptides -
XX isolated using sera from cancer patients, used to develop products
XX for the diagnosis, monitoring or treatment of cancers
XX Claim 67; Page 725; 787pp; English.
XX The invention relates to a method for diagnosing a disorder characterised

CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX Sequence 3127 BP; 1164 A; 500 C; 748 G; 705 T; 10 other;
XX
XX Query Match 76.0%; Score 15.2; DB 20; Length 3127;
XX Best Local Similarity 85.0%; Pred. No. 3.2e+02;
XX Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 1 gagtaggaagatagaacg 20
DB 1002 gagtaggaagatgagagag 1021
||| ||||| ||||| ||
RESULT 14
AAZ39174
ID AAZ39174 standard; DNA; 6774 BP.
XX AAZ39174;
XX 06-MAR-2000 (first entry)
XX Bone morphogenetic protein 4 5' upstream gene sequence with exon 1-3.
XX DE Bone morphogenetic protein 4; BMP-4; promoter; cartilage;
XX KW bone disease; osteohyperplasia; osteometastasis; orthopaedic surgery;
XX KW osteoarthritis; arthroplastis; tumour dissection; bone reconstruction;
XX KW spinal fusion; vertebral canal enlargement; congenital cartilage disease;
XX KW dysosteogenesis; achondroplasia; palatoschisis; mandible reconstruction;
XX KW residual ridge construction; osteoporosis; morphogenesis; hyperplasia;
XX ss.
XX Homo sapiens.
XX WO9957145-A2.
XX 11-NOV-1999.
XX 22-APR-1999; 99WO-IB00732.
XX 30-APR-1998; 98JP-0120173.
XX (HMRI) HOECHST MARION ROUSSEL LTD.
XX Kawai S, Sugitara T;
XX WPI; 2000-062142/05.
XX Human bone morphogenetic protein 4 (BMP-4) promoter used to identify
XX low molecular weight compounds which regulate the expression of BMP-4
XX Claim 1; Page 19-22; 23pp; English.
XX The present sequence represents the 5' upstream gene sequence including
XX the exon 1 through exon 3 regions of human bone morphogenetic protein 4
XX (BMP-4). This region of DNA and a reporter gene can be introduced into
XX a host cell, and used to identify low molecular weight compounds which
XX regulate the expression of human BMP-4. The low molecular weight
XX compounds and their derivatives have morphogenetic activity and
XX inhibiting activity for bone and cartilage through the expression of
XX human BMP-4, and are useful as preventative or therapeutic agents for

CC cartilage and bone diseases, and osteohyperplasia, and as remedies for
 CC osteoarthritis. The compounds are also useful in the fields of
 CC orthopaedic surgery (fracture, osteoarthritis such as joint and hip joint
 CC osteoarthritis, arthrosclerosis, damage of cartilage such as damage of
 CC meniscus, regeneration of bone and cartilage deficit caused by injury
 CC and tumour dissection, bone reconstruction such as spinal fusion and
 CC vertebral canal enlargements, and congenital cartilage and bone diseases
 CC such as dysosteogenesis and achondroplasia); or dental fields (bone
 CC reconstruction such as palatoschisis, mandible reconstruction, and
 CC residual ridge construction) and osteoporosis. The compounds may also
 CC be used for bone graft in aesthetic surgery. The compounds also have
 CC application in the veterinary field. Those compounds that inhibit bone
 CC or cartilage morphogenesis are used for prevention and therapy of bone
 CC and cartilage hyperplasia.

SO Sequence 6774 BP; 1570 A; 1650 C; 1939 G; 1615 T; 0 other;

Query Match 76.0%; Score 15.2; DB 21; Length 6774;
 Best Local Similarity 85.0%; Pred. No. 3.3e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gagtagaggaagtagaagc 20
 ||| ||||| ||||| ||
 Db 3824 gagaggaaggaaggaacg 3843

RESULT 15

AA293815/C
 ID AA293815 standard; DNA; 144460 BP.

AC AA293815;

XX 16-AUG-2000 (first entry)

DE Olfactory receptor operon.

XX Olfactory receptor protein; olfactory; smell; receptor; binding;

KW detection; screening; genotyping; diallelic marker; human; ss.

XX Homo sapiens.

OS Location/Qualifiers

XX Key 2406..2600

FT /tag- a

FT /label- ORF1

FT 9711..10658

FT /tag- b

FT /label- ORF2

FT 24851..25369

FT /tag- c

FT /label- ORF3

FT 45714..46661

FT /tag- d

FT /label- ORF4

FT 80198..81115

FT /tag- e

FT /label- ORF 5

FT 96291..96902

FT /tag- f

FT /label- ORF6

FT 110758..111564

FT /tag- g

FT /label- ORF7

FT 122525..122887

FT /tag- h

FT /label- ORF8

FT 132454..133389

FT /tag- i

FT /label- ORF9

FT 143398..143577

FT /tag- j

FT /label- ORF10

FT misc-feature

FT /tag- k

FT /note- "ubiquitin 1 pseudogene complement"

FT complement (127489..127854)

FT /tag- l

FT /note- "ubiquitin 2 pseudogene complement"

FT 7521

FT /tag- m

FT /note- "Polymorphic base G or T"

FT 8192

FT /tag- n

FT /note- "Polymorphic base G or T"

FT 14483

FT /tag- o

FT /note- "Polymorphic base A or T"

FT 19625

FT /tag- p

FT /note- "Polymorphic base C or T"

FT 20583

FT /tag- q

FT /note- "Polymorphic base A or G"

FT 76947

FT /tag- r

FT /note- "Polymorphic base A or C"

FT 91088

FT /tag- s

FT /note- "Polymorphic base G or C"

FT 91138

FT /tag- t

FT /note- "Polymorphic base A or G"

FT 91187

FT /tag- u

FT /note- "Polymorphic base A or G"

FT 133998

FT /tag- v

FT /note- "Polymorphic base A or G"

FT 140066

FT /tag- w

FT /note- "Polymorphic base C or T"

FT 141176

FT /tag- x

FT /note- "Polymorphic base A or G"

FT 144033

FT /tag- y

FT /note- "Polymorphic base A or G"

FT WO200021985-A2.

FT 20-APR-2000.

FT 13-OCT-1999; 99WO-IB01729.

FT 14-OCT-1998; 98US-0104299.

FT (GEST) GENSET.

FT Bougueleret L, Malekzadeh K;

FT WPI; 2000-317933/27.

FT New nucleic acids encoding ten different olfactory receptor proteins

FT and their diallelic markers, are useful in genetic analysis and in

FT screening for compounds which bind to the receptor proteins

FT Claim 1; Page 103-141; 155pp; English.

FT Ten new olfactory receptor proteins and their diallelic markers have

FT been described. The sequences encoding these receptor proteins and

FT CC which contain the diallelic markers can be used for genotyping. The

FT olfactory receptor proteins can be used to screen for substances

FT which bind to them. See GENSEQ records AA293816-25 and AA293816-95.

FT SO Sequence 144460 BP; 46068 A; 27088 C; 26615 G; 44676 T; 13 other;

Query Match 76.0%; Score 15.2; DB 21; Length 144460;
Best Local Similarity 85.0%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 gactaggaagatagaacg 20
||| ||||| |||||
DB 127449 GAGGAGAGGAGAGAAAG 127430

Search completed: October 16, 2001, 19:01:56
Job time: 3919 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 16, 2001, 19:02:02 ; Search time 119.19 Seconds
(without alignments)
25.531 Million cell updates/sec

Title: US-09-675-650-3

Perfect score: 123

Sequence: 1 MFHISPFKYPHTQKRAQR 23

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP unclassified:*
13: SP vertebrate:*
14: SP virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	41.5	2478	14	09WJC6
2	51	41.5	2485	14	09WJC6
3	51	41.5	2485	14	09WJC6
4	51	41.5	2485	14	09WJC6
5	51	41.5	2485	14	09WJC6
6	51	41.5	2485	14	09WJC6
7	51	41.5	2485	14	09WJC6
8	51	41.5	2485	14	09WJC6
9	51	41.5	2485	14	09WJC6
10	51	41.5	2485	14	09WJC6
11	51	41.5	2485	14	09WJC6
12	51	41.5	2485	14	09WJC6
13	51	41.5	2485	14	09WJC6
14	51	41.5	2485	14	09WJC6
15	51	41.5	2485	14	09WJC6
16	51	41.5	2485	14	09WJC6
17	51	41.5	2485	14	09WJC6
18	51	41.5	2485	14	09WJC6
19	51	41.5	2485	14	09WJC6

20	46	37.4	408	4	09UG14	09UG14 homo sapien
21	46	37.4	602	9	003968	003968 bacterioph
22	46	37.4	1138	2	09PR90	09PR90 ureaplasma
23	46	37.4	1171	5	09VHC2	09VHC2 drosophila
24	46	37.4	2447	14	09WJC6	09WJC6 venezuelan
25	45.5	37.0	349	2	P75821	P75821 escherichia
26	45	36.6	673	10	09SA08	09SA08 arabidopsis
27	44.5	36.2	122	14	06131	06131 vesicular s
28	44.5	36.2	385	13	09IAU0	09IAU0 galus gall
29	44.5	36.2	468	13	073871	073871 galus gall
30	44.5	36.2	474	13	09IBH7	09IBH7 coturnix co
31	44.5	36.2	511	14	088996	088996 vesicular s
32	44.5	36.2	511	14	088999	088999 vesicular s
33	44.5	36.2	511	14	089000	089000 vesicular s
34	44.5	36.2	511	14	089003	089003 vesicular s
35	44.5	36.2	511	14	089007	089007 vesicular s
36	44.5	36.2	511	14	089007	089007 vesicular s
37	44	35.8	119	9	080057	080057 bacterioph
38	44	35.8	291	10	09LME7	09LME7 arabidopsis
39	44	35.8	308	10	043428	043428 daucus caro
40	44	35.8	330	8	09MS01	09MS01 penaea mucr
41	44	35.8	492	5	09NIN9	09NIN9 babesia mic
42	44	35.8	503	5	09NIN8	09NIN8 babesia mic
43	44	35.8	592	5	023013	023013 caenorhabdi
44	44	35.8	746	5	09W4M9	09W4M9 drosophila
45	44	35.8	751	5	097427	097427 drosophila

ALIGNMENTS

RESULT	ID	Query Match	Length	DB ID	Description
1	09WJC6	41.5	2478	14	09WJC6
2	09WJC6	41.5	2485	14	09WJC6
3	09WJC6	41.5	2485	14	09WJC6
4	09WJC6	41.5	2485	14	09WJC6
5	09WJC6	41.5	2485	14	09WJC6
6	09WJC6	41.5	2485	14	09WJC6
7	09WJC6	41.5	2485	14	09WJC6
8	09WJC6	41.5	2485	14	09WJC6
9	09WJC6	41.5	2485	14	09WJC6
10	09WJC6	41.5	2485	14	09WJC6
11	09WJC6	41.5	2485	14	09WJC6
12	09WJC6	41.5	2485	14	09WJC6
13	09WJC6	41.5	2485	14	09WJC6
14	09WJC6	41.5	2485	14	09WJC6
15	09WJC6	41.5	2485	14	09WJC6
16	09WJC6	41.5	2485	14	09WJC6
17	09WJC6	41.5	2485	14	09WJC6
18	09WJC6	41.5	2485	14	09WJC6
19	09WJC6	41.5	2485	14	09WJC6
20	09WJC6	41.5	2485	14	09WJC6
21	09WJC6	41.5	2485	14	09WJC6
22	09WJC6	41.5	2485	14	09WJC6
23	09WJC6	41.5	2485	14	09WJC6
24	09WJC6	41.5	2485	14	09WJC6
25	09WJC6	41.5	2485	14	09WJC6
26	09WJC6	41.5	2485	14	09WJC6
27	09WJC6	41.5	2485	14	09WJC6
28	09WJC6	41.5	2485	14	09WJC6
29	09WJC6	41.5	2485	14	09WJC6
30	09WJC6	41.5	2485	14	09WJC6
31	09WJC6	41.5	2485	14	09WJC6
32	09WJC6	41.5	2485	14	09WJC6
33	09WJC6	41.5	2485	14	09WJC6
34	09WJC6	41.5	2485	14	09WJC6
35	09WJC6	41.5	2485	14	09WJC6
36	09WJC6	41.5	2485	14	09WJC6
37	09WJC6	41.5	2485	14	09WJC6
38	09WJC6	41.5	2485	14	09WJC6
39	09WJC6	41.5	2485	14	09WJC6
40	09WJC6	41.5	2485	14	09WJC6
41	09WJC6	41.5	2485	14	09WJC6
42	09WJC6	41.5	2485	14	09WJC6
43	09WJC6	41.5	2485	14	09WJC6
44	09WJC6	41.5	2485	14	09WJC6
45	09WJC6	41.5	2485	14	09WJC6

Query Match: 41.5%; Score 51; DB 14; Length 2478;
Best Local Similarity: 33.3%; Pred. No. 30;

Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFLHSSPKYPTHOEAKKA 21
:|:::|:| | |:::|
DB 1212 VFINVRTPKYKHHYQCEDHA 1232

RESULT 2

09WJDS PRELIMINARY; PRT; 2485 AA.

AC 09WJDS;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE NONSTRUCTURAL POLYPROTEIN.
OS Venezuelan equine encephalitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZPC738;
RX MEDLINE=99214366; PubMed=10196323;
RA Wang E., Barrera R., Boshell J., Ferro C., Freter J.E., Navarro J.C.,
RA Salas R., Vasquez C., Weaver S.C.;
RT "Genetic and phenotypic changes accompanying the emergence of
RT enzootic subtype IC Venezuelan equine encephalitis viruses from an
RL J. Virol. 73:4266-4271(1999)."

DR EMBL; AF100566; AAD27802.1; -.
DR InterPro; IPR000606; -.
DR InterPro; IPR001788; -.
DR InterPro; IPR002589; -.
DR InterPro; IPR002620; -.
DR Pfam; PF00978; RNA_dep_RNAPol2; 2.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01661; DUF27; 1.
DR Pfam; PF01707; Peptidase_C9; 1.
KW Polyprotein.
SQ SEQUENCE 2485 AA; 276784 MW; 039081920658522 CRC64;

Query Match 41.5%; Score 51; DB 14; Length 2485;
Best Local Similarity 33.3%; Pred. No. 30;
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFLHSSPKYPTHOEAKKA 21
:|:::|:| | |:::|
DB 1212 VFINVRTPKYKHHYQCEDHA 1232

RESULT 3

09JFT5 PRELIMINARY; PRT; 2485 AA.

AC 09JFT5;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE NONSTRUCTURAL POLYPROTEIN.
OS Venezuelan equine encephalitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=243937;
RX MEDLINE=99214366; PubMed=10196323;
RA Wang E., Barrera R., Boshell J., Ferro C., Freter J.E., Navarro J.C.,
RA Salas R., Vasquez C., Weaver S.C.;
RT "Genetic and phenotypic changes accompanying the emergence of
RT enzootic subtype IC Venezuelan equine encephalitis viruses from an
RL J. Virol. 73:4266-4271(1999)."

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=243937;
RA Wang E., Weaver S.C.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF004459; AAD43357.1; -.
DR InterPro; IPR000606; -.
DR InterPro; IPR002589; -.
DR InterPro; IPR002620; -.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01661; DUF27; 1.
DR Pfam; PF01707; Peptidase_C9; 1.
KW Polyprotein.
SQ SEQUENCE 2485 AA; 277022 MW; AACAT28C25F51C07 CRC64;

Query Match 41.5%; Score 51; DB 14; Length 2485;
Best Local Similarity 33.3%; Pred. No. 30;
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFLHSSPKYPTHOEAKKA 21
:|:::~|:~| | |:::~|
DB 1212 VFINVRTPKYKHHYQCEDHA 1232

RESULT 4

09IR88 PRELIMINARY; PRT; 2485 AA.

AC 09IR88;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE NONSTRUCTURAL POLYPROTEIN.
OS Venezuelan equine encephalitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SH3;
RX MEDLINE=99214366; PubMed=10196323;
RA Wang E., Barrera R., Boshell J., Ferro C., Freter J.E., Navarro J.C.,
RA Salas R., Vasquez C., Weaver S.C.;
RT "Genetic and phenotypic changes accompanying the emergence of
RT enzootic subtype IC Venezuelan equine encephalitis viruses from an
RL J. Virol. 73:4266-4271(1999)."

DR EMBL; U55360; AAD43358.1; -.
DR InterPro; IPR000606; -.
DR InterPro; IPR002589; -.
DR InterPro; IPR002620; -.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01661; DUF27; 1.
DR Pfam; PF01707; Peptidase_C9; 1.
KW Polyprotein.
SQ SEQUENCE 2485 AA; 277034 MW; E07E0A68D91140EB CRC64;

Query Match 41.5%; Score 51; DB 14; Length 2485;
Best Local Similarity 33.3%; Pred. No. 30;
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFLHSSPKYPTHOEAKKA 21
:|:::~|:~| | |:::~|
DB 1212 VFINVRTPKYKHHYQCEDHA 1232

RESULT 5

091CG5
ID 091CG5 PRELIMINARY: PRT: 2485 AA.
AC 091R87
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDE.
OS Venezuelan equine encephalitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=66457, AND 66637;
RX MEDLINE=99214366; PubMed=10196623;
RA Wang E., Barrera R., Boshell J., Ferro C., Freier J.E., Navarro J.C.,
RA Salas R., Vasquez C., Weaver S.C.;
RT "Genetic and phenotypic changes accompanying the emergence of
RT epizootic subtype IC Venezuelan equine encephalitis viruses from an
RT enzootic subtype ID progenitor.";
RL J. Virol. 73:4266-4271(1999).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=66457, AND 66637;
RA Wang E., Weaver S.C.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF004472; AAD43356.1; -;
DR EMBL: AF004458; AAD43355.1; -;
DR InterPro: IPR002589; -;
DR InterPro: IPR002589; -;
DR InterPro: IPR002589; -;
DR pfam: PF01661; DUF27.1.
DR pfam: PF01707; Peptidase_C9; 1.
DR Polyprotein.
KW POLYPEPTIDE.
SQ SEQUENCE 2485 AA; 276969 MW; 3F5776AAE8D91F95 CRC64;

Query Match 41.5%; Score 51; DB 14; Length 2485;
Best Local Similarity 33.3%; Pred. No. 30;
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFLHISPFKYPHTOAEOKA 21
Db 1212 VFINVNTPKYHHYQCCEDHA 1232

RESULT 6
OY1R87 PRELIMINARY: PRT: 2492 AA.
AC 091R87
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDE.
OS Venezuelan equine encephalitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=830434;
RX MEDLINE=99214366; PubMed=10196623;
RA Wang E., Barrera R., Boshell J., Ferro C., Freier J.E., Navarro J.C.,
RA Salas R., Vasquez C., Weaver S.C.;
RT "Genetic and phenotypic changes accompanying the emergence of
RT epizootic subtype IC Venezuelan equine encephalitis viruses from an
RT enzootic subtype ID progenitor.";
RL J. Virol. 73:4266-4271(1999).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=830434;
RA Wang E., Weaver S.C.;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: U55362; AAD43359.1; -;
DR InterPro: IPR000606; -;
DR InterPro: IPR002589; -;
DR InterPro: IPR002589; -;
DR pfam: PF01443; Viral_helicase1; 1.
DR pfam: PF01661; DUF27.1.
DR pfam: PF01707; Peptidase_C9; 1.
KW Polyprotein.
SQ SEQUENCE 2492 AA; 277904 MW; E5DE32A5D2D7A3F3 CRC64;

Query Match 41.5%; Score 51; DB 14; Length 2492;
Best Local Similarity 33.3%; Pred. No. 30;
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFLHISPFKYPHTOAEOKA 21
Db 1212 VFINVNTPKYHHYQCCEDHA 1232

RESULT 7
OY2RY3
ID 092RY3 PRELIMINARY: PRT: 740 AA.
AC 092RY3
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NDX2 HOMEBOX PROTEIN (FRAGMENT).
GN NDX2.
OS Lotus japonicus.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Lotus.
OX NCBI_TaxID=34305;
RN [1]
RP SEQUENCE FROM N.A.
RA Jorgensen J.E., Gronlund M., Palisgaard N., Larsen K., Marcker K.A.,
RA Jensen E.;
RT "A new class of plant homeobox genes is expressed in specific regions
RT of determinate symbiotic root nodules.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ011829; CAA09792.1; -;
DR InterPro: IPR001356; -;
DR PROSITE: P550071; HOMEBOX_2; 1.
DR SMART: SM00389; HOX; 1.
KW Homeobox; DNA-binding; Nuclear protein.
FT NON_TER
SQ SEQUENCE 740 AA; 83081 MW; A3D719B8DCFAFC11 CRC64;

Query Match 40.7%; Score 50; DB 10; Length 740;
Best Local Similarity 64.7%; Pred. No. 13;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 3 LHSISPFKYPHTOAEOK 19
Db 438 LFSAGFKONHVOEAK 454

RESULT 8
OYWU1
ID 09WU1 PRELIMINARY: PRT: 1879 AA.
AC 09WU1
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PUTATIVE NONSTRUCTURAL POLYPEPTIDE PRECURSOR.
GN NS.
OS Venezuelan equine encephalitis virus (strain Trinidad donkey).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11038;

SP	SEQUENCE FROM N.A.		
RC	STRAIN-MUCAMBO BEAN 8;		
RX	MEDLINE-99101297; PubMed-9886206;		
RA	Kinney R.M., Pfeiffer M., Tsuchiya K.R., Chang G.J., Roehrig J.T.;		
RT	"Nucleotide sequences of the 26S mRNAs of the viruses defining the		
RT	Venezuelan equine encephalitis antigenic complex.";		
RL	Am. J. Trop. Med. Hyg. 59:952-964(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-MUCAMBO BEAN 8;		
RA	Kinney R.M., Pfeiffer M., Meisner J.;		
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AF075253; AAD14554.1; -		
DR	InterPro: IPR000606; -		
DR	InterPro: IPR002589; -		
DR	InterPro: IPR002620; -		
DR	Pfam: PF01443; Viral_helicase1; 1.		
DR	Pfam: PF01661; DUF27; 1.		
DR	Pfam: PF01707; Peptidase_C9; 1.		
KW	polypotein.		
FT	CHAIN 1 535 METHYLTRANSFERASE NSP1.		
FT	CHAIN 536 1329 NONSTRUCTURAL PROTEINASE NSP2.		
FT	CHAIN 1330 1848 NONSTRUCTURAL PROTEIN NSP3.		
FT	CHAIN 1849 2455 RNA POLYMERASE NS4.		
SQ	SEQUENCE 2455 AA; 273902 MW; 0DDFCB42A809E52 CRR64;		

Query	1	MFHLSPPKTPHTDQAKA	21
	:	:	:
	:	:	:
	:	:	:
	:	:	:
db	1212	VFVNVRTPEYKNNHYQDCEDNA	1232

```

RESULT 10
066594 ID 066594 PRELIMINARY; PRT: 2493 AA.
AC 066594;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NONSTRUCTURAL POLYPROTEIN.
OS Venezuelan equine encephalitis virus (strain TC-83).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11037;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TC-83;
RC MEDLINE=86306669; PubMed=3755750;
RA Johnson B.J., Kliney R.M., Kost C.L., Trent D.W.;
RT "Molecular determinants of alphavirus neurovirulence: nucleotide and
RT deduced protein sequence changes during attenuation of Venezuelan
RT equine encephalitis virus.";
RL J. Gen. Virol. 67:1951-1960(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TC-83;
RC MEDLINE=89243175; PubMed=2524126;
RA Kliney R.M., Johnson B.J., Welch J.B., Tsuchiya K.R., Trent D.W.;
RT "The full-length nucleotide sequences of the virulent Trinidad donkey
RT strain of Venezuelan equine encephalitis virus and its attenuated
RT vaccine derivative, strain TC-83.";
RL Virology 170:19-30(1989).
DR EMBL; L01443; AAR02516.1; -.
DR InterPro; IPR000506; -.
DR InterPro; IPR002589; -.
DR pfam; PF01443; Viral_helicase1; 1.
DR pfam; PF01651; DUF27; 1.
DR pfam; PF01707; peptidase_C9; 1.

```

KW Polypeptide. 1 535 POTENTIAL.
 FT CHAIN 536 1329 POTENTIAL.
 FT CHAIN 1330 1886 POTENTIAL.
 FT CHAIN 1887 2493 POTENTIAL.
 SQ SEQUENCE 2493 AA; 277927 MW; 65388F048664796D CRC64;

Query Match 40.7%; Score 50; DB 14; Length 2493;
 Best Local Similarity 33.3%; Pred. No. 43;
 Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 1 MFLHISPPKYPHTOAEKKA 21
 Db 1212 IFVNWTPKYNHYOCCEDHA 1232

RESULT 11
 ID 090163 PRELIMINARY; PRT: 2493 AA.
 AC 090163;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE NONSTRUCTURAL POLYPEPTIDE.
 OS Venezuelan equine encephalitis virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 OX NCBI_TaxID=11036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-71-180, 600035-71-180/4;
 RX MEDLINE=89243175; PubMed=2524126;
 RA Kinney R.M., Johnson B.J., Welch J.B., Tsuchiya K.R., Trent D.W.;
 RT "The full-length nucleotide sequences of the virulent Trinidad donkey
 strain of Venezuelan equine encephalitis virus and its attenuated
 RT vaccine derivative, strain TC-83.";
 RL Virology 170:19-30(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-71-180, 600035-71-180/4;
 RX MEDLINE=93107872; PubMed=1469368;
 RA Kinney R.M., Tsuchiya K.R., Sneider J.M., Trent D.W.;
 RT "Molecular evidence for the origin of the widespread Venezuelan equine
 encephalitis epizootic of 1969 to 1972.";
 RL J. Gen. Virol. 73:0-0(0).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-71-180, 600035-71-180/4;
 RA Kinney R.M., Trent D.W.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF069903; AAC24033.1; -
 DR InterPro; IPR000606; -
 DR InterPro; IPR002589; -
 DR InterPro; IPR002620; -
 DR Pfam; PF01443; Viral_helicase1; 1.
 DR Pfam; PF01661; DUF27; 1.
 DR Pfam; PF01707; Peptidase_C9; 1.
 KW Polypeptide.
 FT CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.
 FT CHAIN 536 1329 NONSTRUCTURAL PROTEIN NSP2.
 FT CHAIN 1330 1886 NONSTRUCTURAL PROTEIN NSP3.
 FT CHAIN 1887 2493 NONSTRUCTURAL PROTEIN NSP4.
 SQ SEQUENCE 2493 AA; 277876 MW; 9791EE7D2694A466 CRC64;

Query Match 40.7%; Score 50; DB 14; Length 2493;
 Best Local Similarity 33.3%; Pred. No. 43;
 Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 1 MFLHISPPKYPHTOAEKKA 21
 Db 1212 IFVNWTPKYNHYOCCEDHA 1232

RESULT 12
 ID 066592 PRELIMINARY; PRT: 2493 AA.
 AC 066592;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE PUTATIVE NONSTRUCTURAL POLYPEPTIDE PRECURSOR.
 OS Venezuelan equine encephalitis virus (strain Trinidad donkey).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 OX NCBI_TaxID=11038;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TRINIDAD DONKEY;
 RX MEDLINE=86263392; PubMed=3088830;
 RA Kinney R.M., Johnson B.J., Brown V.L., Trent D.W.;
 RT "Nucleotide sequence of the 26 S mRNA of the virulent Trinidad donkey
 strain of Venezuelan equine encephalitis virus and deduced sequence of
 the encoded structural proteins.";
 RL Virology 152:400-413(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TRINIDAD DONKEY;
 RX MEDLINE=86306669; PubMed=3755750;
 RA Johnson B.J., Kinney R.M., Kost C.L., Trent D.W.;
 RT "Molecular determinants of alphavirus neurovirulence: nucleotide and
 RT deduced protein sequence changes during attenuation of Venezuelan
 RT equine encephalitis virus.";
 RL J. Gen. Virol. 67:0-0(0).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TRINIDAD DONKEY;
 RX MEDLINE=89243175; PubMed=2524126;
 RA Kinney R.M., Johnson B.J., Welch J.B., Tsuchiya K.R., Trent D.W.;
 RT "The full-length nucleotide sequences of the virulent Trinidad donkey
 strain of Venezuelan equine encephalitis virus and its attenuated
 RT vaccine derivative, strain TC-83.";
 RL Virology 170:19-30(1989).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TRINIDAD DONKEY;
 RA Kinney R.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; L01442; AAC19321.2; -
 DR InterPro; IPR000606; -
 DR InterPro; IPR002589; -
 DR InterPro; IPR002620; -
 DR Pfam; PF01443; Viral_helicase1; 1.
 DR Pfam; PF01661; DUF27; 1.
 DR Pfam; PF01707; Peptidase_C9; 1.
 KW Polypeptide.
 FT CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.
 FT CHAIN 536 1329 NONSTRUCTURAL PROTEIN NSP2.
 FT CHAIN 1330 1886 NONSTRUCTURAL PROTEIN NSP3.
 FT CHAIN 1887 2493 NONSTRUCTURAL PROTEIN NSP4.
 SQ SEQUENCE 2493 AA; 277929 MW; 7832C85B6D6C0BC CRC64;

Query Match 40.7%; Score 50; DB 14; Length 2493;
 Best Local Similarity 33.3%; Pred. No. 43;
 Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 1 MFLHISPPKYPHTOAEKKA 21
 Db 1212 IFVNWTPKYNHYOCCEDHA 1232

RESULT 13
 Q9WJG9

ID	ORGUC9	PRELIMINARY;	PRT;	2501 AA.
AC	09wtC9			
DT	01-NOV-1999	(TREMBLrel. 12, Created)		
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)		
DT	01-MAR-2001	(TREMBLrel. 16, Last annotation update)		
DE	NONSTRUCTURAL POLYPEPTIDE			
OS	Venezuelan equine encephalitis virus.			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;			
OC	Alphavirus.			
OX	NCBI_TaxID=11036;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-TONATE CAAN 410D;			
RX	MEDLINE=99101297; Pubmed=9886206;			
RA	kinney R.M., Pfeiffer M., Tsuchiya K.R., Chang G.J., Roehrig J.T.;			
RT	"Nucleotide sequences of the 26S mRNAs of the viruses defining the			
RL	Venezuelan equine encephalitis antigenic complex.";			
RL	Am. J. Trop. Med. Hyg. 59:952-964(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-TONATE CAAN 410D;			
RA	kinney R.M., Pfeiffer M., Meisner J.;			
RL	Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AF075254; AADI4556.1; "-			
DR	InterPro; IPR000606; "-			
DR	InterPro; IPR002589; "-			
DR	InterPro; IPR002620; "-			
DR	Pfam; PF01443; Viral_helicase1; 1.			
DR	Pfam; PF01661; DUF27; 1.			
DR	Pfam; PF01707; Peptidase_C9; 1.			
KW	Polyprotein.			
FT	CHAIN	1	METHYLTRANSFERASE NSP1.	
FT	CHAIN	535	NONSTRUCTURAL PROTEINASE NSP2.	
FT	CHAIN	1330	NONSTRUCTURAL PROTEIN NSP3.	
FT	CHAIN	1895	RNA POLYMERASE NSP4.	
QO	SEQUENCE	2501 AA;	278519 MW; 8EAC48DABCF0E28C CRC64;	

Query Match	40.78;	Score 50;	DB 14;	Length 2501;
Best Local Similarity	33.38;	Pred. No. 44;		
Matches	7;	Conservative	8;	Mismatches 6;
				Indels 0;
				Gaps 0;

```
QY      1 MELHISSPFKYRHTQEAQKEA 21
          :|:::|:| | | : : |
Db      1212 VEVNVRTPYKYHHYQGCEDHA 1232
```

RESULT	14			
Q9N8E9				
ID	Q9N8E9	PRELIMINARY;	PRT;	311 AA.
AC	Q9N8E9;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	HYPOTHEICAL 34.7 KDA PROTEIN.			
GN	CHR1.365.			
OS	Trypanosoma brucei.			
OC	Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatida; Trypanosoma			
OX	NCBI_TaxID=5691;			
RP	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=TREU927;			
RA	Hall N., Bowman S., Quail M., Ivens A.C., Kay M.P., Bray-Allen S.,			
RA	Lennard N.J., Clark L.N., Harris B.R., Melville S., Lawson D.,			
RA	Gerrard C., Rajandream M.A., Barrell B.G.;			
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AL359782; CAB95594.1; -.			
KW	Hypothetical protein.			
QO	SEQUENCE 311 AA; 34735 MW; 726802A24459276E CRC64;			

Query Match	38.28;	Score 47;	DB 5;	Length 311;
Best Local Similarity	44.48;	Pred. No. 15;		

Matches	8;	Conservative	4;	Mismatches	6;	Indels	0;	Gaps	0;
QY	6	SSPFKYRHTGEAQKEAQR	23						
		: :							
Db	48	SHPFSHNNHQQQQQQQQR	65						

RESULT 15
Q9N8Q1

DT 01-OCT-2000 (TReMbrel. 15, Created)
 DT 01-OCT-2000 (TReMbrel. 15, Last sequence update)
 DT 01-Mar-2001 (TReMbrel. 16, Last annotation update)
 DE POSSIBLE T16011.22 PROTEIN.
 GN CHR1.202.
 OS Trypanosoma brucei.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma
 OX NCBI_TaxID=5691;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TREU927;
 RA Hall N., Bowman S., Quail M., Ivens A.C., Kay M.P., Bray-Allen S.,
 RA Lennard N.J., Clark L.N., Harris B.R., Melville S., Lawson D.,
 RA Gerrard C., Rastendream M.A., Barrett B.G.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL359782; CAB95482.1; -.
 DR InterPro: IPR001680;
 QO SEQUENCE 1747 AA; 187999 MW; 6E21EEFEEC796824 CRC64;

Query Match	38.2%;	Score 47;	DB 5;	Length 1747;
Best Local Similarity	72.7%;	Pred. No. 89;		
Matches	8;	Conservative	1;	Mismatches 2;
			Indels	0;
			Gaps	0;

```

QY      3 LHISSPEKYPH 13
          |||||: |||
Db     1612 LHISFPYCYPH 1622

```

Search completed: October 16, 2001, 19:07:02
Job time: 300 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 16, 2001, 19:02:37 ; Search time 33.92 Seconds
(without alignments)
23.227 Million cell updates/sec

Title: US-09-675-650-3
Perfect score: 123
Sequence: 1 MFLHSSPFKYPHQEAKKEAQR 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	41.5	2485	1	POLN_EEVV3
2	51	41.5	2485	1	POLN_EEVV3
3	50	40.7	2492	1	POLN_EEVV3
4	46.5	37.8	719	1	TE80_TETTH
5	45.5	37.0	580	1	5NPD_BOOMI
6	44.5	36.2	511	1	VGIG_VSVIG
7	44.5	36.2	511	1	VGIG_VSVIG
8	44.5	36.2	511	1	VGIG_VSVIG
9	44.5	36.2	511	1	VGIG_VSVIG
10	44	35.8	247	1	CEIU_HUMAN
11	44	35.8	297	1	YEPD_YERPE
12	44	35.8	544	1	YI33_CAEEL
13	44	35.8	774	1	SCA_DROME
14	43.5	35.4	258	1	KLKI_PAPHA
15	43.5	35.4	262	1	KLKI_PAPHA
16	43	35.0	628	1	V7OK_TYVY
17	43	35.0	628	1	V7OK_TYVY
18	43	35.0	628	1	V7OK_TYVY
19	42.5	34.6	526	1	MITF_MOUSE
20	42	34.1	201	1	IRB_ARCFU
21	42	34.1	379	1	NIR_PESP
22	42	34.1	395	1	VIBC_VIBCH
23	41.5	33.7	432	1	CGA2_HUMAN
24	41.5	33.7	499	1	CPML_ONCMY
25	41.5	33.7	1135	1	PHYC_SORBI
26	41.5	33.7	1137	1	PHYC_SORBI
27	41	33.3	412	1	ISPL_HUMAN
28	41	33.3	497	1	HMS_DROME
29	41	33.3	905	1	RPA1_THREE
30	40.5	32.9	257	1	KLKI_MACFA
31	40.5	32.9	491	1	CPB4_RABIT
32	40.5	32.9	491	1	CPB5_RABIT
33	40.5	32.9	1005	1	BGAL_ACTPL

34	40	32.5	115	1	YA60_PYRHO
35	40	32.5	129	1	VAL2_ABMVY
36	40	32.5	312	1	YHGX_YEAST
37	40	32.5	339	1	ODBB_PSEPU
38	40	32.5	536	1	C7C2_MAIZE
39	40	32.5	667	1	OS9_HUMAN
40	40	32.5	689	1	TGIC_CAVCU
41	40	32.5	743	1	MYB8_XENLA
42	40	32.5	981	1	MCR_RAT
43	39.5	32.1	491	1	CPB2_RAT
44	39.5	32.1	491	1	CPB2_MOUSE
45	39.5	32.1	500	1	CPBA_MOUSE

ALIGNMENTS

RESULT 1	ID	Query Match	Length	DB ID	Description
1	POLN_EEVV3	STANDARD:	PRT: 2485 AA.		
AC	P36327	01-JUN-1994 (Rel. 29, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	NONSTRUCTURAL POLYPEPTIDE [CONTAINS: NONSTRUCTURAL PROTEINS NSP1 TO NSP4]				
OS	Venezuelan equine encephalitis virus (strain 3880).				
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;				
OC	Alphavirus.				
OX	NCBI_TaxID=36382;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93079659; PubMed=1448915;				
RA	Kimney R.M., Tsuchiya K.R., Snelder J.M., Trent D.W.;				
RT	"Genetic evidence that the epizootic Venezuelan equine encephalitis (VEE) viruses may have evolved from enzootic VEE subtype I-D virus.";				
RL	Virology 191:569-580(1992).				
CC	-1- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.				
CC	-1- FUNCTION: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.				
CC	-1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UGA OCCURS BETWEEN THE CODONS FOR 1871-GLN AND 1872-ARG.				
CC	THIS SWISS-PROT ENTRY IS COPYRIGHTED. IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	EMBL: L00930; AAC19324.1; -				
CC	EMBL: L00930; AAC19326.1; AIT-SEQ.				
CC	PIR: C44213; C44213.				
CC	MEROPS: C09.001; -				
CC	InterPro: IPR000606; -				
CC	InterPro: IPR002589; -				
CC	InterPro: IPR002620; -				
CC	InterPro: DUF27; 1.				
CC	Pfam: PF01661; DUF27; 1.				
CC	Pfam: PF01707; Peptidase-C9; 1.				
CC	Pfam: PF01443; Viral_helicase; 1.				
CC	Polypeptide; Nonstructural protein; RNA-binding; ATP-binding;				
CC	Helicase.				
CC	CHAIN 1 535				
CC	CHAIN 536 1329				
CC	CHAIN 1330 1871				
CC	CHAIN 1872 2485				
CC	NP BIND 721 728				
CC	SEQUENCE 2485 AA; 277282 MW; 83458A964660D639 CRC64;				

Query Match 41.5%; Score 51; DB 1; Length 2485;
Best Local Similarity 33.3%; Prod. No. 9.8;
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

```
OY 1 MFLHSSPKYPTHOAEKKA 21
      :|:::|:|:|:|:|:|:|
Db 1212 IFINWRTPKYHHYQCCEDHA 1232

RESULT 2
POLN_EEVVP ID STANDARD: PRT: 2492 AA.
AC P36328:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDE [CONTAINS: NONSTRUCTURAL PROTEINS NSP1 TO
NSP4].
OS Venezuelan equine encephalitis virus (strain 6676).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
NCBI_TaxID=36385;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93079859; PubMed=1448915;
RA Kliney R.M., Tsuchiya K.R., Snider J.M., Trent D.W.;
RT "Genetic evidence that epizootic Venezuelan equine encephalitis (VEE)
RT viruses may have evolved from enzootic VEE subtype I-D virus.";
RL Virology 191:569-580(1992).
CC -1- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
CC -1- PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UGA OCCURS
CC BETWEEN THE CODONS FOR 1879-GLN AND 1880-ARG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L04653; CAB3722.1; -.
DR PIR: A44213; A44213.
DR MEROPS: C09.001; -.
DR InterPro: IPR000606; -.
DR InterPro: IPR001788; -.
DR InterPro: IPR002589; -.
DR InterPro: IPR002620; -.
DR Pfam: PF01661; DUF27.1.
DR Pfam: PF01707; Peptidase_C9; 1.
DR Pfam: PF00978; RNA_dep_RNAPol2; 2.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Polyprotein; Nonstructural protein; RNA-binding; ATP-binding;
KW Helicase.
FT CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.
FT CHAIN 536 1329 NONSTRUCTURAL PROTEIN NSP2.
FT CHAIN 1330 1879 NONSTRUCTURAL PROTEIN NSP3.
FT CHAIN 1880 2492 NONSTRUCTURAL PROTEIN NSP4.
FT NP_BIND 721 728 ATP (POTENTIAL).
SO SEQUENCE 2492 AA; 277836 MW; E55D056CD7AEDEE CRC64;

Query Match 41.5%; Score 51; DB 1; Length 2492;
Best Local Similarity 33.3%; Pred. No. 9.9;
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFLHSSPKYPTHOAEKKA 21
      :|:::|:|:|:|:|:|:|
Db 1212 IFINWRTPKYHHYQCCEDHA 1232

RESULT 3
POLN_EEVVP ID STANDARD: PRT: 2492 AA.
AC P27282;
```

```
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDE [CONTAINS: NONSTRUCTURAL PROTEINS NSP1 TO
NSP4].
OS Venezuelan equine encephalitis virus (strain Trinidad donkey).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
NCBI_TaxID=11038;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=69243175; PubMed=2524126;
RA Kliney R.M., Johnson B.J.B., Welch J.B., Tsuchiya K.R., Trent D.W.;
RT "The full-length nucleotide sequences of the virulent Trinidad donkey
RT strain of Venezuelan equine encephalitis virus and its attenuated
RT vaccine derivative, strain TC-83.";
RL Virology 170:19-30(1989).
CC -1- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
CC -1- PPM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UGA OCCURS
CC BETWEEN THE CODONS FOR 1879-GLN AND 1880-ARG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04332; AB02518.1; -.
DR PIR: A31467; MNWYTD.
DR MEROPS: C09.001; -.
DR InterPro: IPR000606; -.
DR InterPro: IPR002589; -.
DR InterPro: IPR002620; -.
DR Pfam: PF01661; DUF27.1.
DR Pfam: PF01707; Peptidase_C9; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Polyprotein; Nonstructural protein; RNA-binding; ATP-binding;
KW Helicase.
FT CHAIN 1 535 NONSTRUCTURAL PROTEIN NSP1.
FT CHAIN 536 1329 NONSTRUCTURAL PROTEIN NSP2.
FT CHAIN 1330 1879 NONSTRUCTURAL PROTEIN NSP3.
FT CHAIN 1880 2492 NONSTRUCTURAL PROTEIN NSP4.
FT NP_BIND 721 728 ATP (POTENTIAL).
SO SEQUENCE 2492 AA; 277902 MW; 1BAD415B70DC3FA0 CRC64;

Query Match 40.7%; Score 50; DB 1; Length 2492;
Best Local Similarity 33.3%; Pred. No. 14;
Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

OY 1 MFLHSSPKYPTHOAEKKA 21
      :|:::~|:|:|:|:|:|:|
Db 1212 IFINWRTPKYHHYQCCEDHA 1232

RESULT 4
TE80_TETH ID STANDARD: PRT: 719 AA.
AC Q94818:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE TELOMERASE COMPONENT P80 (EC 2.7.7.-).
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
NCBI_TaxID=5911;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE=95292335; PubMed=7774009;
```

```

RA Collins K., Kobayashi R., Greider C.W.:  
RT "Purification of Tetrahymena telomerase and cloning of genes encoding  
the two protein components of the enzyme."  
RL Cell 81:677-686(1995).  
CC -1- FUNCTION: RIBONUCLEOPROTEIN DNA POLYMERASE THAT CATALYZES THE DE  
NOVO SYNTHESIS OF TELOMERIC SIMPLE SEQUENCE REPEATS. P80 BINDS  
CC TIGHTLY AND SPECIFICALLY TO THE TELOMERASE RNA SUGGESTING ITS  
CC ASSOCIATION WITH A REGION OF RNA SECONDARY STRUCTURE.  
CC -1- SUBUNIT: TELOMERASE CONSIST OF TWO SUBUNIT, P80 AND P95 THAT FORM  
A 1:1 COMPLEX WITH THE 159 NT TELOMERASE RNA.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
DR EMBL; U25641; AAC46601.1; -;  
RW Translase; RNA-directed DNA polymerase; Telomere; Nuclear protein;  
KW DNA-binding.  
SQ SEQUENCE 719 AA; 82351 MW; 8A945A71189CA99C CRC64;
```

Query Match	Best Local Similarity	Matches	Conservative	Mismatches	Indels	Gaps
4	HISSEKYPHTQEA---OKEAOR	23	52.2%	37.0%	52.2%	37.0%
175	YLSPGKVFRTDEACIQREAR	197	52.2%	37.0%	52.2%	37.0%
RESULT 6						
VGIG_VSYIG						
VGIG_VSYIG	STANDARD:	PRT:	511 AA.			
AC	P04883;					
AC	13-AUG-1987 (Rel. 05, Created)					
DT	13-AUG-1987 (Rel. 05, Last sequence update)					
DT	01-OCT-1994 (Rel. 30, Last annotation update)					
DE	SPIKE GLYCOPROTEIN PRECURSOR.					
GN	G.					
OS	Vesicular stomatitis virus (serotype Indiana / strain Glasgow).					
OC	Viruses; ssRNA negative-strand viruses; Mononegavirales;					
OC	Rhabdoviridae; Vesiculovirus.					
OX	NCBI_TaxId=11278;					
RN	[1]					
RF	SEQUENCE FROM N.A.					
RX	MEDLINE=86142630; PubMed=3005478;					
RA	Vandepol S.B., Holland J.J.;					
RT	"Evolution of vesicular stomatitis virus in athymic nude mice: mutations associated with natural killer cell selection.";					
RL	J. Gen. Virol. 67:441-451(1986).					
CC	-1- FUNCTION: THIS PROTEIN FORMS SPIKES ON THE SURFACE OF THE VIRION.					
CC	IT IS RESPONSIBLE BOTH FOR THE BINDING OF THE VIRUS TO SUSCEPTIBLE					
CC	HOST CELLS AND FOR INDUCING THE UPTAKE OF THE VIRUS BY THE CELL.					
CC	THE INTERACTION BETWEEN THE INTERNAL COMPONENTS OF THE VIRION					
CC	AND THE PORTION OF THE GLYCOPROTEIN EXPOSED ON THE CYTOPLASMIC					
CC	FACE OF THE PLASMA MEMBRANE PROBABLY DIRECTS ENVELOPMENT AND					
CC	VIRUS BUDDING.					
CC	-1- PTM: MODIFIED BY THE COVALENT ADDITION OF PALMITIC ACID VIA A					
CC	THIOETHER LINKAGE TO A CYSTEINE.					
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; U08034; AAB38963.1; -					
DR	InterPro; IPR000934; -					
DR	InterPro; IPR002224; -					
DR	PROSITE; PS00785; 5_NUCLEOTIDASE.1; FALSE_NEG.					
DR	PROSITE; PS00786; 5_NUCLEOTIDASE.2; FALSE_NEG.					
DR	Pfam; PF01009; 5_nucleotidase; 1.					
KW	Hydrolase; GPI-anchor; Glycoprotein; signal; zinc.					
FT	NON_TER	1				
FT	SIGNAL	<1 14				
FT	CHAIN	15 552				
FT	PROPEP	553 580				
FT	LIPID	552 580				
FT	CARBOHYD	172 172				
FT	CARBOHYD	172 172				
FT	CARBOHYD	285 285				
FT	CARBOHYD	423 423				
FT	CARBOHYD	536 536				
FT	CONFLICT	15 15				
FT	CONFLICT	37 39				
SO	SEQUENCE	580 AA; 63460 MW; 588EEF2014071AB7 CRC64;				

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X03633; CAA27283.1; -.

DR InterPro: IPR001903; -.

KM Pfam: PF00974; Rhabd_glycop: 1.

CC Transmembrane; Envelope protein; Glycoprotein; Lipoprotein; Signal.

FT SIGNAL 1 16 SPIKE GLYCOPROTEIN.

FT CHAIN 17 511

FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 511 AA; 57691 MW; DDA18842E427C509 CRC64;

Query Match 36.2% Score 44.5; DB 1; Length 511;
 Best Local Similarity 42.9% Pred. No. 20;
 Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 3 LHISSP---FKYPHTQEAQKE 20
 ||||| |::|| | | |
 Db 412 LHLSSKAQVFEPHPIQDASQ 432

RESULT 7
 VGLG_VSVS STANDARD; PRT; 511 AA.

AC P04884;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 01-JUL-1989 (Rel. 11, Last annotation update)
 DE SPIKE GLYCOPROTEIN PRECURSOR.

GN G.

OS Vesicular stomatitis virus (strain Orsay).

OC Viruses; ssRNA negative-strand viruses; Mononegavirales;

CC Rhabdoviridae; Vesiculovirus.

OX NCBI_TaxID=11284;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8518564; PubMed=2985803;
 RA Gallione C.J., Rose J.K.;
 RT "A single amino acid substitution in a hydrophobic domain causes
 RT temperature-sensitive cell-surface transport of a mutant viral
 RT glycoprotein.";

RL J. Virol. 54:374-382(1985).

CC -I- FUNCTION: THIS PROTEIN FORMS SPIKES ON THE SURFACE OF THE VIRION.
 CC IT IS RESPONSIBLE BOTH FOR THE BINDING OF THE VIRUS TO SUSCEPTIBLE
 CC HOST CELLS AND FOR INDUCING THE UPTAKE OF THE VIRUS BY THE CELL.
 CC THE INTERACTION BETWEEN THE INTERNAL COMPONENTS OF THE VIRION
 CC AND THE PORTION OF THE GLYCOPROTEIN EXPOSED ON THE CYTOPLASMIC
 CC FACE OF THE PLASMA MEMBRANE PROBABLY DIRECTS ENVELOPMENT AND
 CC VIRUS BUDDING.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M1048; AAA48438.1; -.

DR InterPro: IPR001903; -.

KM Pfam: PF00974; Rhabd_glycop: 1.

CC Transmembrane; Envelope protein; Glycoprotein; Signal.

FT SIGNAL 1 16

FT CHAIN 17 511 SPIKE GLYCOPROTEIN.

FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARIANT 204 204 F -> S (IN TEMPERATURE-SENSITIVE
 MUTANT TS045, WHICH EXHIBITS TEMPERATURE
 -SENSITIVE CELL-SURFACE TRANSPORT).

SO SEQUENCE 511 AA; 57570 MW; A10E67B6F8138179 CRC64;

Query Match 36.2% Score 44.5; DB 1; Length 511;
 Best Local Similarity 42.9% Pred. No. 20;
 Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 3 LHISSP---FKYPHTQEAQKE 20
 ||||| |::|| | | |
 Db 412 LHLSSKAQVFEPHPIQDASQ 432

RESULT 8
 VGLG_VSVS STANDARD; PRT; 511 AA.

AC P03522;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE SPIKE GLYCOPROTEIN PRECURSOR.

GN G.

OS Vesicular stomatitis virus (strain San Juan).

OC Viruses; ssRNA negative-strand viruses; Mononegavirales;

CC Rhabdoviridae; Vesiculovirus.

OX NCBI_TaxID=11285;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82010868; PubMed=6268840;
 RA Rose J.K., Gallione C.J.;
 RT "Nucleotide sequences of the mRNA's encoding the vesicular stomatitis
 RT virus G and M proteins determined from cDNA clones containing the
 RT complete coding regions.";

RL J. Virol. 39:519-528(1981).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SAN JUAN 56-NM-B;
 RX MEDLINE=90376442; PubMed=2168974;
 RA Bilseil P.A., Nichol S.T.;
 RT "Polymerase errors accumulating during natural evolution of the
 RT glycoprotein gene of vesicular stomatitis virus Indiana serotype
 RT isolates.";

RL J. Virol. 64:4873-4883(1990).

CC -I- FUNCTION: THIS PROTEIN FORMS SPIKES ON THE SURFACE OF THE VIRION.
 CC IT IS RESPONSIBLE BOTH FOR THE BINDING OF THE VIRUS TO SUSCEPTIBLE
 CC HOST CELLS AND FOR INDUCING THE UPTAKE OF THE VIRUS BY THE CELL.
 CC THE INTERACTION BETWEEN THE INTERNAL COMPONENTS OF THE VIRION
 CC AND THE PORTION OF THE GLYCOPROTEIN EXPOSED ON THE CYTOPLASMIC
 CC FACE OF THE PLASMA MEMBRANE PROBABLY DIRECTS ENVELOPMENT AND
 CC VIRUS BUDDING.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: J02428; AAA48370.1; -.

DR M35219; AAA48389.1; -.

DR PIR: A04117; VGVN.

DR InterPro: IPR001903; -.

KM Pfam: PF00974; Rhabd_glycop: 1.

CC Transmembrane; Envelope protein; Glycoprotein; Signal.

FT SIGNAL 1 16

FT CHAIN 17 511 SPIKE GLYCOPROTEIN.

FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 511 AA; 57481 MW; 26E94F713AC0B8E6 CRC64;

Query Match 36.2% Score 44.5; DB 1; Length 511;
 Best Local Similarity 42.9% Pred. No. 20;
 Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

OY 3 LHISSP---FKYPHPOEAOKE 20
 DB 412 LHLSSKAOYFEPHPIODRASQ 432

RESULT 9
 MITE_HUMAN STANDARD: PRT: 526 AA.
 ID MITE_HUMAN 075030: Q14841: Q9P2V0: Q9P2V2: Q9P2V8:
 AC 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE MICROPHthalmia-ASSOCIATED TRANSCRIPTION FACTOR.
 GN MITE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A AND H).
 RC TISSUE=Kidney;
 RX MEDLINE=98321192; PubMed=9647758;
 RA Amoe S., Fuse N., Yasumoto K.-I., Sato S., Tajima I., Yamamoto H.,
 RA Udono T., Duriu Y.K., Tamai M., Takahashi K., Shibahara S.;
 RT "Identification of a novel isoform of microphthalmia-associated
 RT transcription factor that is enriched in retinal pigment
 RT epithelium.";
 RL Biochem. Biophys. Res. Commun. 247:710-715(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM M).
 RC TISSUE=Skin;
 RX MEDLINE=94348499; PubMed=8069297;
 RA Tachibana M., Perez-Jurado L.A., Nakayama A., Hodgkinson C.A., Li X.,
 RA Schneider M., Miki T., Fex J., Francke U., Arneilte H.;
 RT "Cloning of MTF, the human homolog of the mouse microphthalmia gene
 RT and assignment to chromosome 3p14.1-p12.3.";
 RL Hum. Mol. Genet. 3:553-557(1994).
 RN [3]
 RP SEQUENCE OF 1-11 FROM N.A. (ISOFORM M).
 RC TISSUE=Skin;
 RX MEDLINE=98160190; PubMed=9500554;
 RA Watanabe A., Takeda K., Ploplis B., Tachibana M.;
 RT "Epistatic relationship between Waardenburg syndrome genes MTF and
 RT PAX3.";
 RL Nat. Genet. 18:283-286(1998).
 RN [4]
 RP SEQUENCE OF 1-130 FROM N.A. (ISOFORM C).
 RC TISSUE=Kidney;
 RX MEDLINE=20047057; PubMed=10578055;
 RA Fuse N., Yasumoto K.-I., Takeda K., Amoe S., Yoshitawa M., Udono T.,
 RA Takahashi K., Tamai M., Tomita Y., Tachibana M., Shibahara S.;
 RT "Molecular cloning of cDNA encoding a novel microphthalmia-associated
 RT transcription factor isoform with a distinct amino-terminus.";
 RL J. Biochem. 126:1043-1051(1999).
 RN [5]
 RP PARIAL SEQUENCE FROM N.A. (ISOFORMS A; B; H AND M).
 RX MEDLINE=20225461; PubMed=10760582;
 RA Udono T., Yasumoto K.-I., Takeda K., Amoe S., Watanabe K.-I.,
 RA Saito H., Fuse N., Tachibana M., Takahashi K., Tamai M., Shibahara S.;
 RT "Structural organization of the human microphthalmia-associated
 RT transcription factor gene containing four alternative promoters.";
 RL Biochim. Biophys. Acta 1491:205-219(2000).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RC TISSUE=Uterus;
 RX Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
 RT Submitted (Aug-1999) to the EMBL/GenBank/DBD databases.
 RN [7]
 RP MUTAGENESIS, PHOSPHORYLATION SITE SER-405, AND DOMAINS.
 RX MEDLINE=20056131; PubMed=10587587;
 RA Takeda K., Takemoto C., Kobayashi I., Watanabe A., Nobukuni Y.,
 RA Fisher D.E., Tachibana M.;

RT Ser298 of MTF, a mutation site in Waardenburg syndrome type 2, is a
 RT phosphorylation site with functional significance.";
 RL Hum. Mol. Genet. 9:125-132(2000).
 RN [8]
 RP VARIANTS WS2 LYS-310; ARG-324 DEL; PRO-357; ASP-385 AND PRO-405.
 RX MEDLINE=96154685; PubMed=8589691;
 RA Tassabehji M., Newton J.E., Liu X.-Z., Brady A., Donnal D.,
 RA Krajewska-Walasek M., Munday V., Norman A., Oleszczyn E., Reardon W.,
 RA Rice J.C., Trembath R., Wieacker P., Whiteford M., Winter R.,
 RA Read A.P.;
 RT "The mutational spectrum in Waardenburg syndrome.";
 RL Hum. Mol. Genet. 4:2131-2137(1995).
 RN [9]
 RP VARIANT TIEZ SYNDROME LYS-317.
 RC TISSUE=Blood;
 RX MEDLINE=20311381; PubMed=10851256;
 RA Smith S.D., Kelley P.M., Kenyon J.B., Hoover D.;
 RT "Tietz syndrome (hypopigmentation/deafness) caused by mutation of
 RT MTF.";
 RL J. Med. Genet. 37:446-448(2000).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR FOR TYROSINASE AND TYROSINASE-
 CC RELATED PROTEIN 1. BINDS TO A SYMMETRICAL DNA SEQUENCE (E-BOXES)
 CC (5'-CAGCTG-3') FOUND IN THE TYROSINASE PROMOTER. PLAYS A CRITICAL
 CC ROLE IN THE DIFFERENTIATION OF VARIOUS CELL TYPES AS NEURAL CREST-
 CC DERIVED MELANOCYTES, MAST CELLS, OSTEOCLASTS AND OPTIC CUP-DERIVED
 CC RETINAL PIGMENT EPITHELIUM.
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
 CC BHLH PROTEIN. BINDS DNA IN THE FORM OF HOMODIMER OR HETERODIMER
 CC WITH EITHER TP53, TP53 OR TP53.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: 10 ISOFORMS: A1 (SHOWN HERE), A2 B1, B2, C1,
 CC C2, H1, H2, M1 AND M2. ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC THE X2-TYPE ISOFORMS DIFFER FROM THE X1-TYPE BY THE ABSENCE OF
 CC A 6-RESIDUE INSERT.
 CC -1- TISSUE SPECIFICITY: ISOFORM M IS EXCLUSIVELY EXPRESSED IN
 CC MELANOCYTES AND MELANOMA CELLS. ISOFORMS A AND H ARE WIDELY
 CC EXPRESSED IN MANY CELL TYPES INCLUDING MELANOCYTES AND RETINAL
 CC PIGMENT EPITHELIUM (RPE). ISOFORM C IS EXPRESSED IN MANY CELL
 CC TYPES INCLUDING RPE BUT NOT IN MELANOCYTE-LINEAGE CELLS.
 CC -1- PTM: PHOSPHORYLATION AT SER-405 SIGNIFICANTLY ENHANCES THE ABILITY
 CC TO BIND THE TYROSINASE PROMOTER.
 CC -1- DISEASE: DEFECTS IN MTF ARE THE CAUSE OF WAARDENBURG SYNDROME
 CC TYPE 2A (WS2 OR WS2A). A DOMINANT INHERITED DISORDER CHARACTERIZED
 CC BY SENSORINEURAL HEARING LOSS AND PATCHES OF DEPIGMENTATION. THE
 CC FEATURES SHOW VARIABLE EXPRESSION AND PENETRANCE.
 CC -1- DISEASE: DEFECTS IN MTF ARE THE CAUSE OF TIEZ SYNDROME, AN
 CC AUTOSOMAL DOMINANT DISORDER CHARACTERIZED BY GENERALIZED
 CC HYPOPIGMENTATION AND PROFOUND, CONGENITAL, BILATERAL DEAFNESS.
 CC PENETRANCE IS COMPLETE.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. BHLH-ZIP SUBFAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB006909; BAA32288.1; -;
 DR EMBL: AB006989; -; NOT_ANNOTATED_CDS.
 DR EMBL: 229678; CAA82775.1; -;
 DR EMBL: AF034755; AAC39639.1; -;
 DR EMBL: AB032359; BAA95208.1; -;
 DR EMBL: AB032358; BAA95207.1; -;
 DR EMBL: AB032357; BAA95206.1; -;
 DR EMBL: AB009608; BAA95209.1; ALT_TERM.
 DR EMBL: AL10195; CAB53672.1; -;
 DR MIM: 156845; -;
 DR MIM: 103470; -;
 DR MIM: 103500; -;
 DR MIM: 193510; -;

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U50597; AAC46150.1; -

DR InterPro: IPR001626; -

DR Pfam: PF00950; ABC-3; 1.

KW Transport; Transmembrane; Inner membrane.

FT TRANSMEM 20 40 POTENTIAL.

FT TRANSMEM 58 78 POTENTIAL.

FT TRANSMEM 96 116 POTENTIAL.

FT TRANSMEM 133 153 POTENTIAL.

FT TRANSMEM 172 192 POTENTIAL.

FT TRANSMEM 197 217 POTENTIAL.

FT TRANSMEM 224 244 POTENTIAL.

FT TRANSMEM 248 268 POTENTIAL.

SO SEQUENCE 297 AA; 32187 MW; 2D434515C15D46BE CRC64;

Query Match 35.8%; Score 44; DB 1; Length 297;
Best Local Similarity 52.9%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 MFLHSPEKYPHTQEA 17
: | | | | | | | | | |
DB 4 LFLSISEPFAYPFMORA 20

RESULT 12
YL33_CAEEL STANDARD; PRT; 544 AA.

AC P34424;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOTHEICAL 61.8 KDA PROTEIN F44B9.3 IN CHROMOSOME III.
GN F44B9.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berts M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Faveilo A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kersey J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonnenhammer E., Staden K.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: L23648; AAA28033.1; -
DR PIR: S44814; S44814.
DR WormPep; F44B9.3; CE00520.
KW Hypothetical protein.
SO SEQUENCE 544 AA; 61809 MW; 0DE93A8B07A02039 CRC64;

Query Match 35.8%; Score 44; DB 1; Length 544;
Best Local Similarity 58.3%; Pred. No. 25;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 9 FKYPHTQAEKE 20
| | | | | | | | | |
DB 115 FKYPHTQAEKE 126

RESULT 13
SCA_DROME STANDARD; PRT; 774 AA.

AC P21520;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SCABROUS PROTEIN PRECURSOR.
GN SCA.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=91075223; PubMed=2175046;
RA Baker N.E., Mlodzik M., Rubin G.M.;
RT "Spacing differentiation in the developing Drosophila eye: a
RT fibrinogen-related lateral inhibitor encoded by scabrous.";
RL Science 250:1370-1377(1990).
CC -1- FUNCTION: INVOLVED IN REGULATION OF NEUROGENESIS IN DROSOPHILA.
CC SCABROUS MAY ENCODE A LATERAL INHIBITOR OF R8 DIFFERENTIATION.
CC -1- MISCELLANEOUS: POSSESSES FIVE PAIRS OF DIABASIC RESIDUES.
CC -1- SIMILARITY: CONTAINS 1 FIBRINOGEN C-TERMINAL DOMAIN.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: M60065; AAA28880.1; -
DR PIR: A39832; A39832.
DR HSSP; P02671; 1FED.
DR FlyBase: FBgn0003326; sca.
DR InterPro: IPR002181; -
DR Pfam: PF00147; fibrinogen.C.1.
DR PROSITE; PS00514; FIBRIN_Ag_C-DOMAIN.1.
KW Developmental protein; Neurogenesis; Signal.
FT SIGNAL 1 22
FT CHAIN 23 774 SCABROUS PROTEIN.
FT DOMAIN 504 704 FIBRINOGEN C-TERMINAL.
FT DISULFD 517 543 BY SIMILARITY.
FT DISULFD 662 675 BY SIMILARITY.
FT CARBOHYD 347 347 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 562 562 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 635 635 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 719 719 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 762 762 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 774 AA; 87167 MW; 78CD1FB9BA62424P CRC64;

Query Match 35.8%; Score 44; DB 1; Length 774;
Best Local Similarity 41.2%; Pred. No. 36;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 4 HISSPEKYPHTQAEKE 20
| | | | | | | | | |
DB 463 HLNKPKRPHQNVQAO 479

RT with other serine proteases.";
RL Hope-Seyler's Z. Physiol. Chem. 360:1947-1950(1979).
RN [8]
RP SEQUENCE OF 28-47.
RC TISSUE-Urine;
RX MEDLINE-86223893; PubMed-3635530;
RA Takahashi S., Irie A., Katayama Y., Ito K., Miyake Y.;
RT "N-terminal amino acid sequence of human urinary prokallikrein.";
RL J. Biochem. 99:989-992(1986).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
KALLIDIN (LYS-1-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
MET-1-XAA OR LEU-1-XAA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC
DR EMBL; M25629; AAA36136.1; -;
DR EMBL; M33109; AAA59455.1; -;
DR EMBL; M33105; AAA59455.1; JOINED.
DR EMBL; M33106; AAA59455.1; JOINED.
DR EMBL; M33107; AAA59455.1; JOINED.
DR EMBL; M33108; AAA59455.1; JOINED.
DR EMBL; X13561; CAA31912.1; -;
DR EMBL; M12706; AAA59201.1; -;
DR PIR; A24696; KOHU.
DR HSSP; P00757; LSGF.
DR MEROPS; S01.160; -;
DR GLYCOSITE; P06870; -;
DR MIM; 147910; -;
DR InterPro; IPR001254; -;
DR InterPro; IPR001314; -;
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolyase; Serine protease; Kininogenase; Glycoprotein;
KW Multigene family; zymogen; signal.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 262
FT ACT_SITE 65 65
FT ACT_SITE 120 120
FT ACT_SITE 214 214
FT CARBOHYD 93 93
FT CARBOHYD 102 102
FT CARBOHYD 104 104
FT CARBOHYD 108 108
FT CARBOHYD 165 165
FT CARBOHYD 167 167
FT DISULFID 31 174
FT DISULFID 50 66
FT DISULFID 153 220
FT DISULFID 185 199
FT DISULFID 210 235
FT VARIANT 145 145
FT VARIANT 186 186
SQ SEQUENCE 262 AA; 28889 MW; 772AD14A93A496C0 CRC64;
/FTID-VAR_006625.
E -> K.
/FTID-VAR_006626.
FTID-VAR_006626
772AD14A93A496C0 CRC64;
Query Match 35.4%; Score 43.5; DB 1; Length 262;
Best Local Similarity 28.6%; Pred. No. 15;

Matches 8; Conservative 7; Mismatches 4; Indels 9; Gaps 1;
QY 2 FLHISPPFRYP-----HQEAOKE 20
|:|:| |:| |:|:|
DB 89 FVHVSESPHPGFMNLSLEHHTPROADED 116

Search completed: October 16, 2001, 19:07:47
Job time: 310 sec

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	51	41.5	249.2	1	AA4213		nonstructural poly
2	51	41.5	249.2	1	C44213		nonstructural poly
3	50	40.7	249.2	1	MNW9TD		nonstructural poly
4	46.5	37.8	357	2	T29856		probable aspartate
5	46.5	37.8	719	2	S55939		telomerase compone
6	46	37.4	602	2	T13219		major capsid prote
7	46	37.4	1138	2	S68235		membrane nucleas
8	45.5	37.0	349	2	AD4825		hypothetical prote
9	45.5	37.0	349	2	P85598		hypothetical prote
10	44.5	36.6	676	2	B86437		hypothetical prote
11	44.5	36.2	419	2	I38024		Mirf protein - hum
12	44.5	36.2	511	1	GVGN		splice glycoprotein
13	44.5	36.2	520	2	T14752		microphthalma-ass
14	44	35.8	91	2	E86163		protein F15K9.18 [
15	44	35.8	119	2	T00151		hypothetical prote
16	44	35.8	247	2	T17311		hypothetical prote
17	44	35.8	291	2	B86356		hypothetical prote
18	44	35.8	544	2	S44814		hypothetical prote
19	44	35.8	552	2	T25837		F489.3 protein
20	44	35.8	774	2	A39632		hypothetical prote
21	43.5	35.4	262	1	KOHU		scarcous locus (s
22	43	35.0	186	2	B70040		tissue kallikrein
23	43	35.0	239	2	E81697		molybdenum transpo
24	43	35.0	250	2	H85067		ribose 5-phosphat
25	43	35.0	334	2	T23444		hypothetical prote
26	43	35.0	628	2	S01955		hypothetical prote
27	43	35.0	628	2	S19150		hypothetical prote
28	43	35.0	628	2	JQ00110		hypothetical 69k P
29	43	35.0	939	2	T05209		hypothetical prote

30	43	35.0	1069	2	T00043	BH-protocadherin- ϵ
31	43	35.0	1072	2	T00041	BH-protocadherin F
32	43	35.0	1200	2	T00042	BH-protocadherin F
33	43	35.0	2493	3	S26372	nonstructural poly
34	43	35.0	2493	2	S72349	nonstructural poly
35	42.5	34.6	449	2	A40728	microphthalmia-as
36	42.5	34.6	442	2	T27676	hypothetical prote
37	42.5	34.6	1003	2	T19638	hypothetical prote
38	42	34.1	158	2	B85073	hypothetical protei
39	42	34.1	201	1	E69503	indolepyruvate fer
40	42	34.1	379	2	A48936	nitrite reductase
41	42	34.1	335	2	A82283	vibriobactin-speci
42	42	34.1	472	2	G81293	probable type II F
43	42	34.1	619	2	D71361	probable alpha-amyl
44	42	34.1	663	2	T49685	RNA polymerase II
45	42	34.1	843	2	H82362	adenylylate cyclase

nonstructural polyprotein - Venezuelan equine encephalitis virus (strain p676)
N;Contatns: nonstructural protein NS1, nonstructural protein NS2, nonstructural protei
C;Species: Venezuelan equine encephalitis virus
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 17-Feb-1994
C;Accession: A44213
R;Kliney, R.M.; Tsuchiya, K.R.; Sneider, J.M.; Trent, D.W.

R;Klinney, R.M.; Tsuchiya, K.R.; Snelder, J.M.; Trent, D.W

A1>Title: Genetic evidence that epizootic Venezuelan equine encephalitis (VEE) viruses
A1:Reference number: A44213, MUID:93079859

A; Accession: A44213

A;Molecule type: genomic RNA

A;Residues: 1-2492 <K1N>

A:Note: read through the termi
H;C1033-references: GB:L04053

C:superfamily: semliki Forest

C;Keywords: nonstructural pro

F;1-535/Product: nonstructura

F:536-1329/Product: nonstruct

E;1330-1879/Product: nonstruc

E; 1b60-2492/Product: nonstruct

Query Match 41.

Best Local Similarity 33.

Matches 7; Conservatives

QY 1 MELHISSPEKYRHTQEAQK

1213 УЕТНУРКУННУОСЕТ

1212 VE INVALE INI HH100CE

RESULT 2

C44213

nonstructural polyprotein - W

N;Contains: nonstructural, pro

C; Species: Venezuelan equine

C; Date: 17-Feb-1994 #sequence

C, KINNEY, B. M.: Tsuchiya. K. E.
C, KINNEY, B. M.: Tsuchiya. K. E.
C, KINNEY, B. M.: Tsuchiya. K. E.

Virology 191, 569-580, 1992

A; Title: Genetic evidence tha

A;Reference number: A44213; M

A;Accession: C44213

A;Molecule	type:	genomic RNA
1-2403		

A;Residues: 1-2492 <K1N>

A: Note: read through the term

C/Superfamily: Semliki Forest

C) Keywords: nonstructural pro-

File: 1-535/Product: nonstructural

—

A:Accession: F85598
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <STO>
A:Cross-references: GB:AE005174; NID:g12513895; PIDN:AAG55250.1; GSPDB:GN00145; UMGF:
A:Experimental source: strain O157:H7, substrain EDL933
A:Genetics:
A:Gene: Z1102

Query Match 37.0%; Score 45.5; DB 2; Length 349;
Best Local Similarity 40.9%; Pred. No. 19;
Matches 9; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

OY 2 FLHISPP--FKYPHTQEAQKE 20
||||| | : : : :
Db 121 FLHISPPFLYFDYHHNRDKEK 142

RESULT 10
B86437
hypothetical protein AAD21698.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B86437
R:Phenology: A.; Ecker, M.K.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B86437
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-673 <STO>
A:Cross-references: GB:AE005172; NID:g4512629; PIDN:AAD21698.1; GSPDB:GN00141
A:Genetics:
A:Map position: 1

Query Match 36.6%; Score 45; DB 2; Length 673;
Best Local Similarity 41.2%; Pred. No. 45;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 4 HISSPKYPTQEAQKE 20
||||| : : : :
Db 339 HCFKPFEPHLKAAQKE 355

RESULT 11
I38024
MITF protein - human
C:Species: Homo sapiens (man)
C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 05-Nov-1999
C:Accession: I38024
R:TechData, M.; Perez-Durado, L.A.; Nakayama, A.; Hodgkinson, C.A.; Li, X.; Schneidman,
Hum. Mol. Genet. 3, 553-557, 1994
A:Title: Cloning of MITF, the human homolog of the mouse microphthalmia gene and asf1
A:Reference number: I38024; MUID:94348499
A:Accession: I38024
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-419 <RES>
A:Cross-references: EMBL:Z29678; NID:g468496; PIDN:CAA82775.1; PID:g468497
A:Genetics:
A:Gene: GDB:MITF
A:Cross-references: GDB:214776; OMIM:156845; OMIM:193510
A:Map position: 3p14.1-3p12

Query Match	35.2%	Score 44.5;	DB 2;	length 419;
Best Local Similarity	47.1%	Pred. No.33;		
Matches	8;	Conservative	5;	Mismatches
			3;	Indels
				1;
				Gaps
				1;
QY	4	HISPPKPYHTQAOAKE	20	
	11	: : : : :		
Db	15	HLNENPKY-HIQQAQAKQ	30	

RESULT 12
VGVN
antigen protein G precursor - vesicular stomatitis Indiana virus (strain San Juan)

C:Species: Vesicular stomatitis virus
C:Date: 31-Mar-1991 #sequence,revision 19-Feb-1984 #text,change 16-Jul-1999
C:Accession: A04117
R:Rose, J.K.; Gallione, C.J.
J. Virol. 39, 519-528, 1981
A:Title: Nucleotide sequences of the mRNA's encoding the vesicular stomatitis virus G and
A:Reference number: A92983; MUID:82010868
A:Accession: A04117

A:Molecule type: mRNA
A:Residues: 1-511 <ROS>
A:Cross-references: GB:J02428; NID:g335873; PTDN:AAA48370.1; PID:g335877
C:Genetics:
A:Gene: G
C:Superfamily: rhabdovirus spike glycoprotein G
C:Keywords: glycoprotein; spike protein; transmembrane protein
F:1-22/Domain: signal sequence #status predicted <SIS>
F:23-511/Product: spike glycoprotein G #status predicted <SGS>
F:465-481/Domain: transmembrane #status predicted <TMN>
F:119,336/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match	36.2%	Score	44.5	DB	1	Length	511
Best Local Similarity	42.9%	Pred. No.	40				
Matches	9	Conservative	5	Mismatches	4	Indels	3
						Gaps	1

```
OY      3 LHISSP--FKYPHTQEAQKE 20  
         ||::| |::|| |::| :  
Db     412 LHLSKAQVFEPHFIQDAASQ 432
```

RESULT 13
T14752
microphthalmia-associated transcription factor, MTF-A - human
c-Grb10L (Homo sapiens)

Cdate: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
Caccession: T14152. J0207
R:Koehler, K., Beyer, A., Mewes, H.W., Gassenhuber, J., Wiemann, S.
submitted to the Protein Sequence Database, August 1999
#Reference number: Z18180

A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-520 <KOE>
A: Cross-references: EMBL:AL110195
A: Experimental source: adult uterus; clone DKFZ5986B217
R: Name: Fuse, N.; Yasumoto, K.; Sato, S.; Yajima, I.; Yamamoto, H.; Ueno, T. K.; Dutt
Biochem. Biophys. Res. Commun. 247, 710-715, 1998

A:Reference number: JE0207; MUID:98321192
A:Accession: JE0207
A:Molecule type: mRNA
A:Residues: 1-520 <AMA>
A:Cross-references: DDBJ:AB006809; NID:g3413846; PIDD:BAA32288.1; PID:g3413847
C:Comment: This protein plays a important role in differentiation of retinal pigment epithelium.
C:Genetics:
I:Note: DKFZps86b2217.1

```
query match      36.28; Score 44.5; DB 2; Length 520;
```

Best Local Similarity	47.1%	Pred. No. 41;
Matches	8; Conservative	5; Mismatches
		3; Indels
		1; Gaps
		1;
QY	4 HISSPKYPTQEAQKE	20
	1:	
Db	122 HLENPTKY-HIQDAQRQ	137

RESULT 14
E86163
protein F15K9.18 [imported] - Arabidopsis thaliana

C.Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C.Accession: E86163
R.Theologian: A. : Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Denar,
ansen, N.F.; Hughes, B.; Hulzar, L.

A. Authors: Hunter, J. L.; Jenkins, U.; Johnson-Hopson, C.; Kuan, S.; Knaykin, E.; Kuo, A.; Li, J. H.; Li, Y.; Lin, X.; Liu, S.; Liu, Z. A.; Lucero, J. S.; Matti, R.; Marzella, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A. Authors: Salberg, S. L.; Schwartz, J. R.; Shinn, P.; Southwick, A. M.; Sun, H.; Tallonker, M.; Wu, D.; Yu, G.; Fraser, C. M.; Venter, J. C.; Davis, R. W.
Article: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A. Reference number: Ab6141; PMID:21016719

A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-91 <STO>
A.Cross-references: CB:AE005172; NID:G3850586; PIDB:AC72126.1; GSPDB:GN001411
C.Genetics:
A.Gene: F15K9_18
A.Map position: 1

Query Match	35.8%	Score 44	DB 2	Length 91
Best Local Similarity	41.7%	Pred. No. 7.9		
Matches 10; Conservative	3	Mismatches 7	Indels 4	Gaps 1

```

QY      3 LHISSPK-----YPHYQEAQKEAQ 24
      111      1:      111      111:
DB      12 LHIRKVFEDVDPEKHTQNASKEVE 35

```

RESULT 15
T00151
hypothetical protein 18 - Staphylococcus aureus phage phi PVL
C-Saefae; Staphylococcus aureus phage phi PVL

C.Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 21-Jul-2000
C.Accession: T00151
R.Kaneko, Y.; Kimura, T.; Kawakami, Y.; Tomita, T.; Kamio, Y.
Biosci. Biotechnol. Biochem. 61, 1960-1962, 1997
A.Title: Parton-Valentine leuKocJidin genes in a phage-like particle isolated from mlt
A.Reference number: Z14119; MUID:96067870

A: Status: translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-119 <KAN>
A: Cross-references: EMBL:AB009866; NID:g3341907; PIDD:BA31891.1; PID:g3341925
C: Superfamily: Staphylococcus aureus phage phi PVL hypothetical protein 18

Query Match	35.8%	Score 44;	DB 2;	length 119;
Best Local Similarity	47.1%;	Pred. No. 10;		
Matches 8; Conservative	1;	Mismatches 8;	Indels 0;	Gaps 0;

```
QY      2 FLHISSPEKYPHTEAQ 18
          | : | | | | |
Db     103 FAYFDGPKLPKTEVQ 119
```

Search completed: October 16, 2001, 19:04:50

Job time: 207 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 16, 2001, 18:30:52 ; Search time 53.97 Seconds
(without alignments)
8.775 Million cell updates/sec

Title: US-09-675-650-3
Perfect score: 123
Sequence: 1 MFHISPPKRYPTQEAQKRAQR 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46.5	37.8	699	US-08-851-843A-52	Sequence 52, Appl
2	46.5	37.8	699	US-08-974-549A-188	Sequence 188, Appl
3	46.5	37.8	719	US-08-851-843A-7	Sequence 7, Appl
4	46.5	37.8	719	US-08-974-549A-219	Sequence 219, Appl
5	44.5	36.2	511	US-08-646-695-5	Sequence 5, Appl
6	44.5	36.2	511	PCR-US96-06053-5	Sequence 5, Appl
7	44	35.8	492	US-08-845-258-21	Sequence 21, Appl
8	44	35.8	492	US-08-990-571-21	Sequence 21, Appl
9	44	35.8	503	US-08-845-258-52	Sequence 52, Appl
10	44	35.8	503	US-08-990-571-52	Sequence 52, Appl
11	43.5	35.4	238	US-08-944-483-39	Sequence 39, Appl
12	43.5	35.4	258	US-08-744-026-3	Sequence 3, Appl
13	43.5	35.4	258	US-09-102-732-3	Sequence 3, Appl
14	43.5	35.4	258	US-09-261-767-3	Sequence 3, Appl
15	43.5	35.4	262	US-08-744-026-4	Sequence 4, Appl
16	43.5	35.4	262	US-08-790-137-1	Sequence 1, Appl
17	43.5	35.4	262	US-08-790-137-3	Sequence 3, Appl
18	43.5	35.4	262	US-08-681-151-4	Sequence 4, Appl
19	43.5	35.4	262	US-09-102-732-4	Sequence 4, Appl
20	43.5	35.4	262	US-08-824-874-4	Sequence 4, Appl
21	43.5	35.4	262	US-08-807-151-4	Sequence 4, Appl
22	43.5	35.4	262	US-09-261-767-4	Sequence 4, Appl
23	43.5	35.4	262	US-09-210-084-4	Sequence 4, Appl
24	43	35.0	747	US-09-035-648-18	Sequence 18, Appl
25	43	35.0	747	US-09-001-951-18	Sequence 18, Appl
26	42	34.1	127	US-08-853-659A-57	Sequence 57, Appl
27	41.5	33.7	432	US-08-522-166-8	Sequence 8, Appl

28	41.5	33.7	432	1	US-08-488-382A-8	Sequence 8, Appl
29	41.5	33.7	432	2	US-08-480-912-8	Sequence 8, Appl
30	40	32.5	301	5	PCT-US95-13975-72	Sequence 72, Appl
31	40	32.5	338	4	US-08-403-545-3	Sequence 3, Appl
32	40	32.5	338	4	US-08-404-381-3	Sequence 3, Appl
33	40	32.5	435	3	US-08-911-321-8	Sequence 2, Appl
34	39.5	32.1	1239	2	US-08-937-931-2	Sequence 2, Appl
35	39.5	32.1	1239	4	US-09-285-502-2	Sequence 2, Appl
36	39	31.7	40	3	US-08-926-842B-51	Sequence 51, Appl
37	39	31.7	313	3	US-08-926-842B-62	Sequence 62, Appl
38	38	30.9	488	1	US-07-794-393-2	Sequence 2, Appl
39	38	30.9	488	1	US-08-001-711-2	Sequence 2, Appl
40	38	30.9	488	3	US-08-704-711A-22	Sequence 22, Appl
41	38	30.9	489	4	US-08-448-488-11	Sequence 11, Appl
42	38	30.9	532	3	US-08-481-435-12	Sequence 12, Appl
43	38	30.9	553	3	US-08-481-435-11	Sequence 11, Appl
44	38	30.9	566	2	US-08-272-255-8	Sequence 8, Appl
45	38	30.9	566	5	PCT-US95-08565-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-851-843A-52
Sequence 52, Application US/08851843A
Patent No. 6093809
GENERAL INFORMATION:
APPLICANT: Cecch, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morin, Gregg B.
APPLICANT: Harley, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6093809e1 Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,843A
FILING DATE: 06-MAY-1997
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-00293005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 52:

```

: SEQUENCE CHARACTERISTICS:
:     LENGTH: 659 amino acids
:     TYPE: amino acid
:     STRANDEDNESS:
:     TOPOLOGY: linear
:     MOLECULE TYPE: peptide
: US-08-851-843A-52

```

Query Match	37.8%	Score 46.5	DB 3	length 699
Best Local Similarity	41.4%	Pred. No. 9.7		
Matches 12	Conservative 1	Mismatches 7	Indels 9	Gaps 1

```

QY      2 FLHISSP-----FKYPHTQEAQKEA 21
          | | | | | | | | | |
Db      280 FCHISEPKERVYKILGKKYPTKEEYKAA 308

```

RESULT 2
US-08-974-549A-188
Sequence 188, Application US/08974549A

GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morlin, Gregg B.
APPLICANT: Hatley, Calvin B.
APPLICANT: Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:

```

1  APPLICATION NUMBER:  WO PCT/US97/17618
2  FILING DATE:  01-OCT-1997
3  PRIOR APPLICATION DATA:
4  APPLICATION NUMBER:  WO PCT/US97/17885
5  FILING DATE:  01-OCT-1997
6  ATTORNEY/AGENT INFORMATION:
7  NAME:  Apple, Randolph Ted
8  REGISTRATION NUMBER:  36,429
9  REFERENCE/DOCKET NUMBER:  015389-00261005US
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE:  (415) 576-0200
12 TELEFAX:  (415) 576-0300
13 INFORMATION FOR SEQ ID NO:  188:
14 SEQUENCE CHARACTERISTICS:
15     LENGTH:  699 amino acids
16     TYPE:  amino acid
17     STRANDEDNESS:
18     TOPOLOGY:  linear
19     MOLECULE TYPE:  peptide
20     US-08-974-549A-188

```

Query Match:	37.8%	Score	46.5	DB	4	Length	699
Best Local Similarity:	41.4%	Pred. No.	9.7				
Matches	12	Conservative	1	Mismatches	7	Indels	9
						Gaps	1

```
QY      2 FLHISSP-----FKYPHTQEAOKKA 21
          |   ||| |           ||| : | | |
Db     280 FCHISEPKERVKYLGGKKYPKTEEEYKAA 308
```

RESULT 3
US-08-851-843A-7
; Sequence 7, Application US/08851843A
; Patent No. 6003800

GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
APPLICANT: Lingner, Joachim
APPLICANT: Nakamura, Toru
APPLICANT: Chapman, Karen B.
APPLICANT: Morlin, Gregg B.
APPLICANT: Harley, Calvin
APPLICANT: Andrews, William H.
TITLE OF INVENTION: No. 6093809e1 Telomerase
NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,843A
FILING DATE: 06-MAY-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
CLASSIFICATION:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	5
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	---

LENGTH: 511 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-646-695-5

Query Match 36.2%; Score 44.5; DB 4; Length 511;
Best Local Similarity 42.9%; Pred. No. 14;
Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

OY 3 LHSSP---FKYPTQEOAKE 20
||:| | :| | | :| | :
Db 412 LHSSKAQVFEHPHQDASQ 432

RESULT 6
PCT-US96-06053-5
Sequence 5, Application PC/TUS9606053
GENERAL INFORMATION:
APPLICANT: Yale University
TITLE OF INVENTION: RECOMBINANT VESICULOVIRUSES AND THEIR
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/06053
FILING DATE: 01-MAY-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6523-009-228
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 669-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 511 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-06053-5

Query Match 36.2%; Score 44.5; DB 5; Length 511;
Best Local Similarity 42.9%; Pred. No. 14;
Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

OY 3 LHSSP---FKYPTQEOAKE 20
||:| | :| | | :| | :
Db 412 LHSSKAQVFEHPHQDASQ 432

RESULT 7
US-08-845-258-21
Sequence 21, Application US/08845258
Patent No. 6183976
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
ATTORNEY/AGENT INFORMATION:
NAME: Lodes, Michael J.

APPLICANT: Houghton, Raymond
APPLICANT: Sleath, Paul R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/845,258
FILING DATE: 24-APR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.426C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-845-258-21

Query Match 35.8%; Score 44; DB 4; Length 492;
Best Local Similarity 31.8%; Pred. No. 17;
Matches 7; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

OY 2 FLHSSPKYPTQEOAKEQR 23
| :| | :| | | :| | :
Db 103 FLVSKVEYEHTEHLAKCHK 124

RESULT 8
US-08-990-571-21
Sequence 21, Application US/08990571
Patent No. 6214971
GENERAL INFORMATION:
APPLICANT: Reed, Steven G. et al.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B.
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/990,571
FILING DATE: 11-DEC-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392

? REFERENCE/DOCKET NUMBER: 210121.42662
?
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (206) 622-4900
? TELEFAX: (206)682-6031
? INFORMATION FOR SEQ ID NO: 21:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 492 amino acids
? type: amino acid
? STRANDEDNESS:
? TOPOLOGY: linear
? NS-08-990-571-21

Query Match	35.88;	Score 44;	DB 4;	Length 492;
Best Local Similarity	31.88;	Pred. No. 17;		
Matches	7;	Conservative	8;	Mismatches 7;
				Indels 0;
				Gaps 0;

```
OY      2 FLHISSPKYPHTQEAQKEAQR 23
          | :| | ::| | :| ::
Db     103 FRYISKEYEYEHTELAKHECK 124
```

RESULT 9
US-08-845-258-52
; Sequence 52, Application US/08845258
; Patent No. 6183976

1 APPLICANT: Reed, Steven G.
 2 APPLICANT: Lodes, Michael J.
 3 APPLICANT: Houghton, Raymond
 4 APPLICANT: Sleath, Paul R.
 5 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
 6 TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
 7 NUMBER OF SEQUENCES: 53
 8
 9 CORRESPONDENCE ADDRESS:
 10 ADDRESSEE: SEED AND BERRY
 11 STREET: 6300 Columbia Center, 701 Fifth Avenue
 12 CITY: Seattle
 13 STATE: Washington
 14 COUNTRY: USA
 15 RX: 009104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/845,258
FILING DATE: 24-Apr-1997

? ATTORNEY/AGENT INFORMATION:
 ? NAME: Maki, David J.
 ? REGISTRATION NUMBER: 31,392
 ? REFERENCE/DOCKET NUMBER: 210121.426C1
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (206) 622-4900
 ? TELEFAX: (206) 682-6031
 ? INFORMATION FOR SEO ID NO: 52

LENGTH: 503 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Babesia Microti
 ?
 ?
 US-08-845-258-52

Query Match	35.8%;	Score 44;	DB 4;	Length 503;
Best Local Similarity	31.8%;	Pred. No. 17;		
Matches	7;	Conservative	8;	Mismatches 7;
				Indels 0;
				Gaps 0;

Qy	2	FLHISSPKYPHTQEAQKEAQR	23
		: : : :	:
Db	31	FRYISKEYEYEHTELAKHECKK	52

RESULT 10
 US-081-990-571-52
 Sequence 52, Application us/0890571
 Patent No. 6214971
 GENERAL INFORMATION:
 APPLICANT: Reed, Steven G. et al.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B
 NUMBER OF SEQUENCES: 79
 CORRESPONDENCE ADDRESS:
 ADDRESSSEE: steed, and, berry

ADDRESS: 6300 Columbia Center, 701 Fifth Avenue
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/990,571

1 FILING DATE: 11-DEC-1997
2 CLASSIFICATION:
3 ATTORNEY/AGENT INFORMATION:
4 NAME: MAKI, DAVID J.
5 REGISTRATION NUMBER: 31,392
6 REFERENCE/DOCKET NUMBER: 210121.426C2
7 TELECOMMUNICATION INFORMATION:
8 TELEPHONE: (206) 622-4900
9 TELEFAX: (206)682-6031
10 INFORMATION FOR SEQ ID NO: 52:

SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Babesia Microti
DS-08-990-571-52

Query Match	35.8%	Score 44	DB 4	Length 503
Best Local Similarity	31.8%	Pred. No. 17		
Matches 7, Conservative	8	Mismatches 7	Indels 0	Gaps 0

Qy	2	FLHISSPKYPHTQEAQKEAQR	23
	1	: : : :	:
Db	31	FRYISKEYEYETELAKEHCKK	52

RESULT 11
US-08-944-483-39
; Sequence 39, Application US/08944483
; Patent No. 623246
; GENERAL INFORMATION:

APPLICANT: COHEN, MAURICE
APPLICANT: COLPITTS, TRACEY L.
APPLICANT: FRIEDMAN, PAULA N.
APPLICANT: GRAMADOS, EDWARD N.
APPLICANT: KLASS, MICHAEL R.
APPLICANT: RUSSELL, JOHN C.
APPLICANT: STEWART, KENT D.
APPLICANT: STROUPE, STEVEN D.
TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
OF THE PROSTATE

```

NUMBER OF SEQUENCES: 76
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/944,483
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Becker, Cheryl L.
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6183. US. 01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO. 6232456e
US-08-944-483-39

```

PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0154 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-835-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 258 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GenBank
 CLONE: 871B14
 US-08-744-026-3

;
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 871814
; US-09-102-732-3

Query Match	35.4%	Score 43.5;	DB 2;	length 258;
Best Local Similarity	28.6%	Pred. No. 9.3;		
Matches	8;	Conservative	7;	Mismatches 4;
				Indels 9;
				Gaps 1

```

QY      2 FLHISSPEKYP-----HTQEAOKE 20
          |::| |::|
Db      86 FVHVSESPHPCFNMSLKNHTRQADED 113

```

```

14 RESULT 14
15 US-09-261-767-3
16 ; Sequence 3, Application US/09261767
17 ; Patent No. 6184357
18 ; GENERAL INFORMATION:
19 ; APPLICANT: Bandman, Olga
20 ; APPLICANT: Goll, Surya K.
21 ; TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC
22 ; TITLE OF INVENTION: KALIKREIN
23 ; NUMBER OF SEQUENCES: 5
24 ; CORRESPONDENCE ADDRESS:
25 ; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
26 ; STREET: 3174 Porter Drive
27 ; City: Palo Alto
28 ; STATE: CA
29 ; COUNTRY: US
30 ; Zip: 94304
31 ; COMPUTER READABLE FORM:
32 ; MEDIUM TYPE: Diskette
33 ; COMPUTER: IBM compatible
34 ; OPERATING SYSTEM: DOS
35 ; SOFTWARE: FASTSEQ Version 2.0
36 ; CURRENT APPLICATION DATA:
37 ; APPLICATION NUMBER: US/09/261,767
38 ; FILING DATE:
39 ; CLASSIFICATION:
40 ; PRIOR APPLICATION DATA:
41 ; APPLICATION NUMBER: 08/744,026
42 ; FILING DATE:
43 ; ATTORNEY/AGENT INFORMATION:
44 ; NAME: Billings, Lucy J
45 ; REGISTRATION NUMBER: 36,749
46 ; REFERENCE/DOCKET NUMBER: PF-0154 US
47 ; TELECOMMUNICATION INFORMATION:
48 ; TELEPHONE: 415-855-0555
49 ; TELEFAX: 415-845-4166
50 ; INFORMATION FOR SEQ ID NO: 3:
51 ; SEQUENCE CHARACTERISTICS:
52 ; LENGTH: 258 amino acids
53 ; TYPE: amino acid
54 ; STRANDEDNESS: single
55 ; TOPOLOGY: linear
56 ; IMMEDIATE SOURCE:
57 ; LIBRARY: GenBank
58 ; CLONE: 871814
59 ;
60 ; US-09-261-767-3

```

Query Match	35.48;	Score 43.5;	DB 4;	Length 258;
Best Local Similarity	28.68;	Pred. No. 9.3;		
Matches	8;	Conservative	7;	Mismatches 4;
				Indels 9;
				Gaps 1

```

OY 2 FLHISSPFKYP-----HTQEAQKE 20
      |::| |::|
Db 86 FVHVSESPHPCEFMSLKNHTROADED 113

```

RESULT 15

US-08-744-026-4
: Sequence A Application US/08744026

Query Match	35.48;	Score 43.5;	DB 1;	length 262;
Best Local Similarity	28.68;	Pred. No. 9.5;		
Matches	8;	Conservative	7;	Mismatches 4;
				Indels 9;
				Gaps 1;

QY	2	FLHISSPEKYP-----HTQEAQKE	20
		:: ::	
Db	89	FVHVSESPHPGFNMSLLENHTROADED	116

Search completed: October 16, 2001, 19:03:38
Job time: 1966 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 17:49:27 ; Search time 1981.17 Seconds
(without alignments)
2414.303 Million cell updates/sec

Title: US-09-675-650-1

Perfect score: 506
Sequence: 1 cagggaagcacaagaagc.....ggcttacaagaacatgcaac 506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: em_gss_fun:*
182: em_gss_hum1:*
183: em_gss_hum2:*
184: em_gss_hum3:*
185: em_gss_hum4:*
186: em_gss_hum5:*
187: em_gss_hum6:*
188: em_gss_hum7:*
189: em_gss_hum8:*
190: em_gss_hum9:*
191: em_gss_inv1:*
192: em_gss_inv2:*
193: em_gss_inv3:*
194: em_gss_inv4:*
195: em_gss_inv5:*
196: em_gss_inv6:*
197: em_gss_inv7:*
198: em_gss_inv8:*
199: em_gss_inv9:*
200: em_gss_inv10:*
201: em_gss_inv11:*
202: em_gss_inv12:*
203: em_gss_inv13:*
204: em_gss_inv14:*
205: em_gss_inv15:*
206: em_gss_inv16:*
207: em_gss_inv17:*
208: em_gss_inv18:*
209: em_gss_inv19:*
210: em_gss_inv20:*
211: em_gss_inv21:*
212: em_gss_inv22:*
213: em_gss_inv23:*
214: em_gss_inv24:*
215: em_gss_inv25:*
216: em_gss_inv26:*
217: em_gss_inv27:*
218: em_gss_inv28:*
219: em_gss_inv29:*
220: em_gss_inv30:*
221: em_gss_inv31:*
222: em_gss_inv32:*
223: em_gss_inv33:*
224: em_gss_inv34:*
225: em_gss_inv35:*
226: em_gss_inv36:*
227: em_gss_inv37:*
228: em_gss_inv38:*
229: em_gss_inv39:*
230: em_gss_inv40:*
231: em_gss_inv41:*
232: em_gss_inv42:*
233: em_gss_inv43:*
234: em_gss_inv44:*
235: em_gss_inv45:*
236: em_gss_inv46:*
237: em_gss_inv47:*
238: em_gss_inv48:*
239: em_gss_inv49:*
240: em_gss_inv50:*
241: em_gss_inv51:*
242: em_gss_inv52:*
243: em_gss_inv53:*
244: em_gss_inv54:*
245: em_gss_inv55:*
246: em_gss_inv56:*
247: em_gss_inv57:*
248: em_gss_inv58:*
249: em_gss_inv59:*
250: em_gss_inv60:*
251: em_gss_inv61:*
252: em_gss_inv62:*
253: em_gss_inv63:*
254: em_gss_inv64:*
255: em_gss_inv65:*
256: em_gss_inv66:*
257: em_gss_inv67:*
258: em_gss_inv68:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

TITLE	JOURNAL MEDLINE	COMMENT
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)	20202663
Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br		
This sequence was derived from the FAPSP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/genhtml2.pl?IL1-IL2&t2=IL2-FT0159-070800-120-H01&t3=2000-08-07&t4=1) Seq primer: puc 18 forward High quality sequence stop: 167.		
FEATURES		
SOURCE	1..167 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_id="FT0159" /dev_stage="Adult" /note="Organ: prostate,tumor; Vector: puc18; Site:1: Smal; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."	
BASE COUNT	47 a 38 c 49 g 33 t	
ORIGIN		
Query Match	28.1%; Score 142.4; DB 147; Length 167;	
Best Local Similarity	99.3%; Pred. No. 6.9e-31;	
Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	250 caccagatccctgggagaagaatcccgccgcacatcttgggtatgatgaagctgcgacctg 309 	
Db	24 CAGAGATCCCGGGGAGGAATGCCCGCCGCATCTTGGGTATCGATAGCCCTCCCTG 83	
QY	310 tgcctgggtcccgctgttagggaagacatagaagaatgaattgattgttccctaagaag 369 	
Db	84 TGCCGTGTCGCCGCTTGTTGTAGGGAAGACATTAGAAAATGCAATTGATGTGTCTTTAAAGG 143	
QY	370 atgggcaggaanaacagatcctgtt 393 	
Db	144 ATGGCAGGAAAACAGATCGTGT 167	
RESULT 2		
LOCUS	AA578773	
DEFINITION	nl24a04.s1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:953262, mRNA sequence.	
ACCESSION	AA578773	
VERSION	AA578773.1	
KEYWORDS	GI:2356957	
SOURCE	EST.	
ORGANISM	human.	
REFERENCE	Homo sapiens	
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
TITLE	1 (bases 1 to 402)	
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps-jemail.nih.gov Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui, M.D.	

[illegible][illegible]

RESULT	11	BE054713	780 bp	mRNA	EST	07-MAR-2001
LOCUS	11	BE054713/c				
DEFINITION		GA_Ea0034J02f Gossypium arboreum 7-10 dpa fiber library Gossypium				
ACCESSION		BE054713				
VERSION		BE054713				
KEYWORDS		BE054713.2 GI:13246558				
SOURCE		EST				
ORGANISM		Gossypium arboreum.				
REFERENCE		Gossypium arboreum.				
AUTHORS		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.				
TITLE		1 (bases 1 to 780)				
JOURNAL		Wing, R.A., Fisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry D., Wood, T.C., Leslie, A. and Wilkins, T.A.				
COMMENT		An integrated analysis of the genetics, development, and evolution of the cotton fiber unpublished (2000)				
FEATURES		On Jun 8, 2000 this sequence version replaced gi:8381770.				
Source		Contact: Wing RA				
		Clemson University Genomics Institute				
		100 Jordan Hall, Clemson, SC 29634, USA				
		Tel: 864 656 7288				
		Fax: 864 656 4293				
		Email: rwing@clemson.edu				
		Seq primer: TAAATGACCTCATATGAGG				
		High quality sequence stop: 256.				
		Location/Qualifiers				
		1..780				
		/organism="Gossypium arboreum"				
		/strain="AKA"				
		/cultivar="8400"				
		/db_xref="taxon:29729"				
		/clone="GA_Ea0034J02f"				
		/clone_lib="Gossypium arboreum 7-10 dpa fiber library"				
		/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"				
		/lab_host="E. coli"				
		/note="vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI"				
BASE COUNT		261 a 111 c 300 g 108 t				
ORIGIN						
Query Match		7.7%; Score 39; DB 162; Length 780;				
Best Local Similarity		52.8%; Pred. No. 1.3;				
Matches	84;	Conservative 0; Mismatches 75; Indels 0; Gaps 0				
Yr	65	tttttgatggccttaagtcctcctactcgtttctatcctctcctactcaactgtctccgg 124				
Db	497	ttcttttttcccttccttccttccttccttccttccttccttccttccttccttccttc 438				
Yr	125	aatcacactacagatcttctatctctgtcctgctgattgtctgactggtcactcgattta 184				
Db	437	cctccggcttccttccttccttccttccttccttccttccttccttccttccttccttc 378				
Yr	185	tctctacgagatctgattttcttaccgggactcctc 223				
Db	377	ccctcccccgggtttgcttttcccttccttccttccttccttccttccttccttccttc 339				
RESULT	12					
ID	BF266315/c					
AC	BF266315	standard; RNA; EST; 1231 BP.				
SV	BF266315.1					
DT	20-NOV-2000 (rel. 65, Created)					

[illegible]

[illegible]

	FEATURES	Source
	collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/The Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp. the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .	/organism="Drosophila melanogaster" Location/Ovaliflers 1. 1031 /db_xref="taxon:7227" /clone_lib="RPCL-98" /clone="BACR25K05" /note="end : TET3"
BASE COUNT	176 a 169 c 124 g 214 t 348 others.	
ORIGIN		
QUERY Match	7.7%; Score 38.8; DB 219; Length 1031;	
Best Local Similarity	25.3%; Pred. No. 1.6;	
Matches	76; Conservative 85; Mismatches 139; Indels 0; Gaps 0;	
OY	11 aaagggaagccagaggaagtgccttataagaacccaattactaccagaattttg 70	
Dd	989 AAAAAAAMAAAMAMAMCAAMAMMAHAMAMATCTCTCMTMWTATMWTMYTC 930	
OY	71 atggcctaagtcctctaccgcgttctcatctctccctaccatcgctcccgaatoca 130	
Dd	929 HTTAHTATAATMYHMACAGCATYTMTTWTWTCWTAAGTCGCHNTHTTTTHMAATPAMHA 870	
OY	131 ctaccagattctcatcttctcctcgatgatcttgtaactgtagctcaattgattacctoa 190	
Dd	869 CTCMCNHYMYMHWTCTAHATATCTMCSCYYMTTWMMHCCSTAAHTYTMHNHTTYTWT 810	
OY	191 cgagdgctcgattctctaccgccggcgctcaaccgcctccctccatatattgctcccacttc 250	
Dd	809 AMNMATYTMACMCMTMAWCYHCMAAAAYMMWMYMWYCMMCCAAATMTWTTCTMMACCTCTM 750	
OY	251 acagatccccctggagaaatcgccggccgcacatcttggtgcatcgatgagctcgccctgt 310	
Dd	749 HCCMCMACCTMYMYTCYHMTMYCTGCCCTCCMYYYYAHHTMAHMAHMAACCATATCTATAT 690	
RESULT 15		
AO895392/c	1049 bp DNA GSS 10-NOV-1999	
LOCUS	HS-3143-B2.C10-T7C CIT Approved Human Genomic Sperm Library D Homo	
DEFINITION	sapiens genomic clone Plate=3143 Col=20 Row=F, DNA sequence.	
ACCESSION	AO895392	
VERSION	AO895392.1 GI:6351582	
KEYWORDS	GSS.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1049)	
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.	
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome	
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)	
MEDLINE	99380589	
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington	

401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: <http://www.hnsc.washington.edu>
 Plate: 3143 row: F column: 20
 Seq primer: T7
 Class: BAC ends

High quality sequence stop: 1049.

FEATURES

source

Location/Qualifiers

1. .1049
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-3143 Col-20 Row-F"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelOBAC1; BAC clones in
 E-Coli DH10B"

BASE COUNT 347 a 32 c 572 g 55 t 43 others

ORIGIN

Query Match 7.7%; Score 38.8; DB 235; Length 1049;

Best Local Similarity 51.1%; Pred. No. 1.6;

Matches 91; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

OY 83 tccctactggtctatcactctcactcactgctcccggaatccactacgatttc 142
 |||||
 DB 771 TACCCTCCTCCTTTCTCTCTCTTCACCCCTCCCTCTCTCTCCCTCAATTCAC 712
 OY 143 tatctctgctcgtatgtctgactgctcactgattatccctcagagatcgat 202
 |||||
 DB 711 TTTTCTCTCCTTTTCCTCCCTCCCTCCCTCCCTCCCTCTTCTCTCTCTCTCT 652
 OY 203 ttctaccgggctcactcctcgtccctccatattgtcctcactttcacagatccct 260
 |||||
 DB 651 TCTCCTTCTCCTCCCTCCCTCTTCTCCCTCCCTCCCTCTCTCTCTCTCTCTCT 594

Search completed: October 16, 2001, 18:30:40

Job time: 2473 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 17:53:02 ; Search time 88.04 Seconds
(Without alignments) 1088.044 Million cell updates/sec

Title: US-09-675-650-1
Perfect score: 506
Sequence: 1 ccggaagcacaagaagc.....ggttcacaagacatgcaac 506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCYUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.2	11.5	7218	1	US-08-232-463-14
2	35	6.9	10207	1	US-08-920-812-2
3	35	6.9	10207	1	US-08-920-812-2
4	35	6.9	10207	1	US-08-921-177-2
5	35	6.9	10207	1	US-08-362-577C-2
6	35	6.9	10207	1	US-08-920-828-2
7	33.2	6.6	15231	3	US-09-128-155-16
8	33	6.5	7218	1	US-08-232-463-14
9	32.6	6.4	46899	1	US-08-471-119A-1
10	31.6	6.2	176373	3	US-09-128-155-17
11	30.6	6.0	4758	3	US-09-191-647-1
12	29.8	5.9	53526	3	US-08-658-136-2
13	29.8	5.9	53577	3	US-08-658-136-1
14	29.6	5.8	4177	3	US-09-023-082A-23
15	29.4	5.8	2085	2	US-08-668-128B-7
16	29.4	5.8	2085	2	US-08-905-445-7
17	29.2	5.8	1838	5	PCT-US93-06251-85
18	29.2	5.8	4175	1	US-08-306-691B-49
19	29.2	5.8	4175	4	US-08-202-841A-9
20	29.2	5.8	4175	5	PCT-US93-06251-84
21	29.2	5.8	7785	2	US-08-276-967-1
22	29.2	5.8	15231	3	US-09-128-155-16
23	29	5.7	4517	4	US-09-140-804-9
24	28.8	5.7	1022	4	US-08-960-780-33
25	28.8	5.7	1022	4	US-09-073-898-33
26	28.8	5.7	1952	1	US-08-333-358-1
27	28.8	5.7	1952	1	US-08-463-694-1

28	28.8	5.7	1952	1	US-08-694-501-1	Sequence 1, Appl
29	28.8	5.7	3460	2	US-08-751-305-1	Sequence 1, Appl
30	28.8	5.7	4074	1	US-08-471-033-19	Sequence 19, Appl
31	28.8	5.7	4074	2	US-08-471-044-19	Sequence 19, Appl
32	28.8	5.7	4074	2	US-08-463-483A-19	Sequence 19, Appl
33	28.8	5.7	4074	2	US-08-471-046A-19	Sequence 19, Appl
34	28.8	5.7	4074	2	US-08-470-566B-19	Sequence 19, Appl
35	28.8	5.7	4074	2	US-08-469-334-19	Sequence 19, Appl
36	28.8	5.7	4074	3	US-09-300-529-19	Sequence 19, Appl
37	28.8	5.7	9370	1	US-08-320-559-27	Sequence 27, Appl
38	28.8	5.7	9370	3	US-08-545-860D-27	Sequence 27, Appl
39	28.8	5.7	9370	5	PCT-US94-04496-27	Sequence 27, Appl
40	28.8	5.7	9391	1	US-08-320-559-25	Sequence 25, Appl
41	28.8	5.7	9391	3	US-08-545-860D-25	Sequence 25, Appl
42	28.8	5.7	9391	5	PCT-US94-04496-25	Sequence 25, Appl
43	28.6	5.7	301	2	US-08-332-766A-23	Sequence 23, Appl
44	28.4	5.6	291	1	US-07-922-723A-7	Sequence 7, Appl
45	28.4	5.6	291	1	US-07-799-828C-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOXLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14

[illegible]

RESULT 2
 US-08-920-812-2/c
 Sequence 2, Application US/08920812
 Patent No. 5763188
 GENERAL INFORMATION:
 APPLICANT: Ohno, Tsuneya
 APPLICANT: Matsuhisa, Akio
 APPLICANT: Uehara, Hirotsugu
 APPLICANT: Eda, Soji
 TITLE OF INVENTION: Probe for Diagnosing Infectious Diseases
 NUMBER OF SEQUENCES: 25
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 STREET: 6300 Sears Tower, 233 South Wacker Drive
 City: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606-6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/920,812
 FILING DATE: 29-AUG-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/362,577
 FILING DATE: 27-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Rin-Laures, Li-Hsien
 REGISTRATION NUMBER: 33,547
 REFERENCE/DOCKET NUMBER: 19036/32420
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312/474-6300
 TELEFAX: 312/474-0448
 TELEX: 25-3856
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 10207 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 ORIGINAL SOURCE:
 ORGANISM: Staphylococcus aureus
 STRAIN: Clinical isolate SA-24

Query Match	6.9%	Score 35;	DB 1;	Length 10207;
Best Local Similarity	47.9%;	Pred. No. 0.2;		
Matches 101; Conservative	0;	Mismatches 110;	Indels 0;	Gaps 0.
Oy 283	cttggtcatcatgatagcctgcgcctctgtgcctctgtcccgctgtgagggagacattag	342		
Db 2910	CATCAGCTCTTCATATAGCGGTACGATTCGATTTCACCTTGGCCATTTAGATTCATTA	2851		
Oy 343	aaatgaattgatgtgtcctctaagaagatgggcaggaanaacagatcctgttgatatt	402		
Db 2850	TCAATGAACGAAATCATTTTCCATGACCGCGATGATGTGTATCGAACGCTTGTATCAATGT	2791		
Oy 403	tatttgaacgggattacagatttgaatgaagtccaccaaagtgcagcataccaatggagag	462		
Db 2790	TATATGCAATATTATTAAGCCGCTAATATGCAATCAATTAATTAACATTAAGACTCGTCAC	2731		
Oy 463	gaaacagagcagaaatccttgatgtctca	493		
Db 2730	GATATTAATTTGCATATCTTGATATTCGTAA	2700		

```

3
US-08-920-827-2/c
; Sequence 2, Application US/08920827
; Patent No. 5770375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hitotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rln-Laures, Ll-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: Clinical Isolate SA-24
US-08-920-827-2

```

Query Match	5.9%	Score 35	DB 1	Length 10207
Best Local Similarity	47.9%	Pred. 0.2	Mismatches 110	Indels 0
Matches 101	Conservative 0	Mismatches 110	Indels 0	Gaps 0
QY 283	cttgggtcatcgtatgagcctcgccctgtgcctgtgcccgtctgtgaggaagacattag	342		
Db 2910	CATCGTCTTCATATGAGGGTAFGCATTGGATTCCATTGCCAATTGATTAATCAATCTCA	2851		
QY 343	aaaagaattgattggttcctctaagaagatggcagaagaaacgatacccgctgttgataatt	402		
Db 2850	TCAATGAACGATCAATCTTCATGACCCGGATGTGTGTCTTCCAAAGCTTTGTATCATGT	2791		
QY 403	tattgaacgagatcacagaattgaaatgaagtcaccaagtgagacatcaccaatgagag	462		
Db 2790	TATAAGAAATATTATAAAGCGGTATATTGGAATCAATAATAAATTAACATTAAGACTGTCAC	2731		
QY 463	gaaacagacgaanaatccttgatgcttca	493		
Db 2730	GATATAAATTGCATATCTTGATATTCGTA	2700		

```

RESULT 4
US-08-921-177-2/c
: Sequence 2, Application US/08921177
: Patent No. 5798211
: GENERAL INFORMATION:
: APPLICANT: Ohno, Tsuneya
: APPLICANT: Matsuhisa, Akio
: APPLICANT: Uehara, Hirotsugu
: APPLICANT: Eda, Soji
: TITLE OF INVENTION: Probe for Diagnosing Infectious Diseases
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/921,177
: FILING DATE: 29-AUG-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/362,577
: FILING DATE: 27-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Rin-Laures, Li-Hsien
: REGISTRATION NUMBER: 33,547
: REFERENCE/DOCKET NUMBER: 19036/32420
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312/474-0448
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10207 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: ORIGINAL SOURCE:
: ORGANISM: Staphylococcus aureus
: STRAIN: Clinical Isolate SA-24
US-08-921-177-2

```

	Query Match	Similarity	6.9%	Score 35	DB 1	Length 10207	
	Best Local	Similarity	47.9%	Pred. No. 0.2			
	Matches	101	Conservative	0	Mismatches	110	Indels 0; Gaps 0;
OY	283	cttggatcagatgagcctgcgccttgctcgtgctgtgccgcttgyagggagacattag	342				
DB	2910	CATCGCTTTCATTAAGCGGTACGATTCGATTCACCTTGCCAAATTAATGAATCAACTTCA	2851				
OY	343	aaaatgaatttgatgcttccttaagaagatggcgagaaacagatccctgttgatatt	402				
DB	2850	TCAATGAACGATTCATTTCATGACCGGATGATGTGTATCAACGCTTTGTATCAATCT	2791				
OY	403	tatttgaacggattacagatttgaattgaattgaccacaagatgagcaattaccatgagag	462				
DB	2790	TATAAGGATATATTATAAAGCGGCTAATATGGAATCAAAATTAATAGCATAGACTGTCAC	2731				
OY	463	gaaacagacgagaaacttgatgctttca	493				
DB	2730	GATATTAATTTGCCAATATCTTGATATTCGAA	2700				

```

US-08-1362-577C-2/c
; Sequence 2, Application US/08362577C
; Patent No. 5807673
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,577C
; FILING DATE: 27-MAR-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: Clinical Isolate SA-24
; US-08-362-577C-2

```



```

; ADDRESSER: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)863-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpt-F15
; US-08-232-463-14

Query Match
Best Local Similarity 6.5%; Score 33; DB 1; Length 7218;
Matches 6; Conservative 105; Mismatches 60; Indels 0; Gaps 0;

QY 327 gaggaagacattagaataatgatgtgtctcttaagatggcaggaacaga 366
DB 1229 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1170
QY 387 tccgtgtgatatattattgaacggattacagattgaatgaagtcacaaatga 446
DB 1169 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1110
QY 447 gcattccaatgaggaagaaacagacagagaatacttgatgcttcaagaag 497
DB 1109 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1059

RESULT 9
US-08-471-119A-1
; Sequence 1, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergerdorfer, Kurt
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
```

```

; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONF/CONF
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
; US-08-471-119A-1

Query Match
Best Local Similarity 6.4%; Score 32.6; DB 1; Length 46899;
Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 185 tctcaagagatcgtgattctaccggggtacacctgcctccatattgtctc 244
DB 38506 TTCTGCCCGGCTCTAGAGATGCTCCCGGATCCCTCGTTCAACATCATTTGTGTC 38565
QY 245 attccacagatccctggagaaatgc 271
DB 38566 ACTCCAGCGAATCTCGTAACTGC 38592

RESULT 10
US-09-128-155-17
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(176373)
; OTHER INFORMATION: n = A,T,C or G
; US-09-128-155-17
```

```

Query Match          6.2%: Score 31.6; DB 3; Length 176373;
Best Local Similarity 50.7%: Pred. No. 14;
Matches 76; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

OY 82 tctctactacgtgtctctacctctctactacactgtctctcccggaatccactacgatttt 141
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122325 tctcttctcgtctgtctgtctctctctctctctctctctctctctctctctctt 122384

OY 142 ctatttctctcgtctgtatgtctgacgtgctcacttgattatctctcaaggagctgga 201
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122385 tctcttctctctctctctctctctctctctctctctctctctctctctctctt 122444

OY 202 ttcttaaccgggtccactcgcgtccccc 231
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122445 tctacttctctctctctctctctctctctctc 122474

RESULT 11
US-09-191-647-1
: Sequence 1, Application US/09191647
: Patent No. 6046015
: GENERAL INFORMATION:
: APPLICANT: Goodman, Corey
: APPLICANT: Kid, Thomas
: APPLICANT: Brose, Katja
: APPLICANT: Tessier-Lavigne, Marc
: TITLE OF INVENTION: Modulating Robo: Ligand Interactions
: FILE REFERENCE: B98-031-3
: CURRENT APPLICATION NUMBER: US/09/191, 647
: CURRENT FILING DATE: 1998-11-13
: EARLIER APPLICATION NUMBER: 60/065,544
: EARLIER FILING DATE: 1997-11-14
: EARLIER APPLICATION NUMBER: 60/081,057
: EARLIER FILING DATE: 1998-04-07
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
: LENGTH: 4758
: TYPE: DNA
: ORGANISM: human
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(4575)
US-09-191-647-1

Query Match          6.0%: Score 30.6; DB 3; Length 4758;
Best Local Similarity 53.8%: Pred. No. 3.7;
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 279 ccaatcttggtcatcgcagacgcctcgcctgtgctgtgtcccgctgtgtgaggaaggaca 338
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2824 ccaatgtgttcgaaggcgcagacgtgtgagtcctcaatcatcgatcgatcaaccaca 2883

OY 339 ttgaagaatgatgatgtgtgtctctctaagaatggtgcaggaataacagatccctgtgt 395
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2884 tgaatacatgtaggaactgtccacttaagaagaagaagaagatgatctctgtgt 2940

RESULT 12
US-08-658-136-2
: Sequence 2, Application US/08658136
: Patent No. 607117
: GENERAL INFORMATION:
: APPLICANT: KLINGER, KATHERINE W
: APPLICANT: LANDES, GREGORY M
: APPLICANT: BURN, TIMOTHY C
: APPLICANT: CONNORS, TIMOTHY D
: APPLICANT: DACKOWSKI, WILLIAM
: APPLICANT: GERMINO, GREGORY
: APPLICANT: QIAN, FENG
: TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
: NUMBER OF SEQUENCES: 58

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: GENA-17,8
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-872-8400
TELEFAX: 508-872-5415
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 53526 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-658-136-2

Query Match      5.9%; Score 29.8; DB 3; Length 53526;
Best Local Similarity 48.0%; Pred. No. 28;
Matches 85; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy      83   tccctactggtttcattaccccttcaccacactgcgtcccggaatcaccaagatttc 142
Db       TCCCCTCCCTCCCTTTCGTGTTCTCTTCCTTCCCTCCACACTTCCCTTCCTTCCC 35842
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy      143   tattcttgcctcgtatgtctgacctgacagtgcacttgatattatacccaggagtcgat 202
Db     35843 TCTCCTTCTCTCTTCCTTCCCTGCCTGCCCTCTCGTCTTCCTTCCTTCCTCC 35902
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy      203   ttctaccgggcatcacctccgltccctccatatattgtcctccaatttacagatccc 259
Db     35903 CCTCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 35959
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-08-658-136-1
; Sequence 1, Application US/08658136
; Patent No. 6071717
GENERAL INFORMATION:
APPLICANT: KLINGER, KATHERINE W
APPLICANT: LANDES, GREGORY M
APPLICANT: BURN, TIMOTHY C
APPLICANT: CONNORS, TIMOTHY D
APPLICANT: DACIKOWSKI, WILLIAM
APPLICANT: GERMINO, GREGORY
APPLICANT: QIAN, FENG
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: ONE MOUNTAIN ROAD
CITY: FRAMINGHAM
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 17:56:37 ; Search time 164.45 Seconds
(without alignments)
1932.005 Million cell updates/sec

Title: US-09-675-650-1
Perfect score: 506
Sequence: 1 caggaagcacaagaaggaac.....gcttcacaagacatgcaac 506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDS8/gcgdata/geneseq/geneseq/NA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseq/NA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseq/NA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseq/NA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseq/NA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseq/NA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseq/NA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseq/NA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseq/NA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseq/NA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseq/NA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseq/NA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseq/NA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseq/NA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseq/NA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseq/NA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseq/NA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseq/NA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseq/NA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseq/NA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseq/NA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	243.4	48.1	812	21	Human immunogenic
2	243.4	48.1	820	19	Prostate cancer an
3	243.4	48.1	1872	19	Prostate cancer an
4	243.4	48.1	2037	19	Prostate cancer an
5	243.4	48.1	2229	21	Human immunogenic
6	243.4	48.1	2426	21	Human immunogenic
7	243.4	48.1	3112	21	Human immunogenic
8	243.4	48.1	3582	19	Prostate cancer an
9	224.6	44.4	597	20	Human secreted pro
10	218.4	43.2	718	21	Human immunogenic
11	138	27.3	437	21	Human secreted pro

c	12	61	12.1	301	21	AA06520	Human immunogenic
c	13	52.4	10.4	359	20	AA233445	Human prostate can
c	14	37.6	7.4	1515	22	AAE59232	Human PTH2 recepto
c	15	35.8	7.1	936	22	AAE58252	Oligonucleotide D1
c	16	35.8	7.1	936	22	AAE58254	Oligonucleotide D1
c	17	35.8	7.1	936	22	AAE58257	Oligonucleotide D1
c	18	35.8	7.1	936	22	AAE58259	Oligonucleotide D2
c	19	35.8	7.1	936	22	AAE58262	Oligonucleotide D2
c	20	35.8	7.1	938	22	AAE58255	Oligonucleotide D1
c	21	35.4	7.0	122186	22	AAE58256	Human histone deac
c	22	35	6.9	753	18	AAE84068	DNA encoding a S.
c	23	35	6.9	936	22	AAE58252	Oligonucleotide D1
c	24	35	6.9	936	22	AAE58254	Oligonucleotide D1
c	25	35	6.9	936	22	AAE58257	Oligonucleotide D2
c	26	35	6.9	936	22	AAE58259	Oligonucleotide D2
c	27	35	6.9	936	22	AAE58262	Oligonucleotide D1
c	28	35	6.9	938	22	AAE58255	Oligonucleotide D1
c	29	35	6.9	10207	15	AA05134	Staphylococcus aur
c	30	34.8	6.9	982	20	AAE37417	Human secreted pro
c	31	34.6	6.8	133719	21	AAE64754	Macaca mulatta rha
c	32	34.4	6.8	119211	22	AAE28553	Genomic fragment #
c	33	33.4	6.6	4594	18	AAV74511	Staphylococcus aur
c	34	33.2	6.6	3264	21	AAE43195	Arabidopsis thalia
c	35	33.2	6.6	3333	21	AAE49503	Arabidopsis thalia
c	36	33.2	6.6	11901	20	AAE02998	Human IL-1ra BAC c
c	37	32.8	6.5	365	21	AAE22045	Human secreted pro
c	38	32.6	6.4	46899	15	AAE54386	T. niveum Cyclospo
c	39	32.4	6.4	578	21	AAE45319	Human secreted exp
c	40	32.4	6.4	6608	18	AAE42751	Chicken CHD-1A gen
c	41	32.2	6.4	277	21	AAE00950	Human colon cancer
c	42	32.2	6.4	2427	21	AAE95570	Nucleic acid sequ
c	43	31.6	6.3	847	21	AAE34356	Arabidopsis thalia
c	44	31.6	6.2	348	20	AAE51622	Human secreted pro
c	45	31.6	6.2	693	20	AAE97741	Extended human sec

ALIGNMENTS

RESULT	1	
AAA06690/c	AAA06690 standard; cDNA; 812 BP.	
11	AAA06690;	
AC	13-JUN-2000 (first entry)	
XX	Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:471.	
XX	Human; prostate cancer; diagnosis; tumour; gene therapy; detection;	
XX	Immunogenic; cytostatic; vaccine; ss.	
OS	Homo sapiens.	
XX	NO200004149-A2.	
PD	27-JAN-2000.	
XX	14-JUL-1999;	99MO-US15838.
XX	14-JUL-1998;	98US-0115453.
XX	14-JUL-1998;	98US-0116134.
PR	23-SEP-1998;	98US-0159812.
PR	23-SEP-1998;	98US-0159822.
PR	15-JAN-1999;	99US-0232149.
PR	15-JAN-1999;	99US-0232880.
PR	09-APR-1999;	99US-0288946.
XX	(CORI-) CORIXA CORP.	
PA	Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;	
XX	WPI, 2000-171268/15.	

XX New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 262; 263pp; English.
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumor protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AA06241 to AA06691 and
CC AA08200 to AA08202 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 812 BP; 198 A; 189 C; 169 G; 256 T; 0 other;
Query Match 48.1%; Score 243.4; DB 21; Length 812;
Best Local Similarity 99.2%; Pred. No. 7.9e-70;
Matches 255; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 250 caccgattccctggggaagaatgccggccgcatcttggtatgatgagctgcgctg 309
DB 541 CAGAGATCCCTGGGGAATGCCGGCCGATCTGGGTATCGATGAGCTCGCCTG 482
QY 310 tgcctgtcccgctgtgaggaagacatagaataatgatgtgttccttaag 369
DB 481 TGCCTGTCCCGCTGTGTAGGGAAGACATTAAGAAATGATGTATGCTTCTTAAGG 422
QY 370 atgggcaggaagaacagatcctgtgtgatatatttgaaacgggattacgattgaa 429
DB 421 ATGGGCAGGAAGAACAAGATCCTGTGTGATTTATTGAAACGGGATTAACGATTTGAAA 362
QY 430 tgaagtcacaaagtgaagcatlaccaatgagaagaaacagaacgagaataatcttgatgac 489
DB 361 TGAAGTCA-CAAAGTGACATTACCAATGAGAGGAAACAGACGAGAAATCTTGATGGC 303
QY 490 ttcaagaacatgcaac 506
DB 302 TTCACAAGACATGCAAC 286
RESULT 2
AAV62429 standard; cDNA; 820 BP.
XX ID AAV62429 standard; cDNA; 820 BP.
XX AAV62429;
AC Homo sapiens.
XX OS
XX PN MO9845420-A1.
XX PD 15-OCT-1998.
XX PF 09-APR-1998; 98MO-CA00346.
XX PR 10-APR-1997; 97US-0041836.
XX (DIAG-) DIAGNOCURE INC.
XX PA
XX

PI Bussemakers MJG;
XX
DR WPI; 1998-568347/48.
XX
PT New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
XX prevention and treatment of prostatic cancer
PS Claim 4; Pages 77-78; 111pp; English.
CC The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 3 sequence comprising of exons 1, 3, and 4a
CC of the PCA3 gene. The PCA3 cDNA splice variant 3 sequence,
CC isolated from a human primary prostatic tumor tissue cDNA library,
CC was found in approximately 15% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunoassay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumour. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (ant)agonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX
SQ Sequence 820 BP; 262 A; 169 C; 191 G; 198 T; 0 other;
Query Match 48.1%; Score 243.4; DB 19; Length 820;
Best Local Similarity 99.2%; Pred. No. 7.9e-70;
Matches 255; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 250 caccgattccctggggaagaatgccggccgcatcttggtatcatgtagctgcgctg 309
DB 277 caggatccctggggaagaatgccggccgcatcttggtatcatgtagctgcgctg 336
QY 310 tgcctgtcccgctgtgaggaagacatagaataatgatgtgttccttaag 369
DB 337 tgcctgtcccgctgtgaggaagacatagaataatgatgtgttccttaag 396
QY 370 atgggcaggaagaacagatcctgtgtgatatatttgaaacgggattacagattgaa 429
DB 397 atgggcaggaagaacagatcctgtgtgatatatttgaaacgggattacagattgaa 456
QY 430 tgaagtcacaaagtgaagcatlaccaatgagaagaaacagaacgagaataatcttgatgac 489
DB 457 tgaagtcacaaagtgaagcatlaccaatgagaagaaacagaacgagaataatcttgatgac 515
QY 490 ttcaagaacatgcaac 506
DB 516 ttcaagaacatgcaac 532
RESULT 3
AAV62428 standard; cDNA; 1872 BP.
XX ID AAV62428 standard; cDNA; 1872 BP.
XX AAV62428;
AC Homo sapiens.
XX OS
XX PN 30-DEC-1998 (first entry)
XX DT
XX DE Prostate cancer antigen (PCA3) cDNA splice variant 2.
XX DE Prostate cancer antigen (PCA3) cDNA splice variant 2.
XX DE Prostate cancer antigen cDNA splice variant 2; PCA3; prostatic cancer;
XX PC; ds.
XX OS
XX PA Homo sapiens.
XX

PM W09845420-A1.
XX 15-OCT-1998.
PD 09-APR-1998; 98WO-CA00346.
XX 10-APR-1997; 97US-0041836.
XX (DIAG-) DIAGNOCURE INC.
XX Bussemakers MDG;
XX WPI, 1998-568347/48.
XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT prevention and treatment of prostatic cancer
PS Claim 4; Pages 76-77; 111pp; English.
XX The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 2 sequence comprising of exons 1, 3, 4a and
CC 4b of the PCA3 gene. The PCA3 cDNA splice variant 2 sequence,
CC isolated from a human primary prostatic tumor tissue cDNA library,
CC was found in approximately 65% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunosay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumor. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (antagonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX Sequence 1872 BP; 567 A; 389 C; 369 G; 539 T; 8 other;
SQ

Query Match 48.1%; Score 243.4; DB 19; Length 1872;
Best Local Similarity 99.2%; Pred. No. 1.2e-69;
Matches 255; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 250 cacagatccctggagaatgcgcgcgcacatcttggtcatcatgagctgcgcctg 309
DB 277 cagagatccctggagaatgcgcgcgcacatcttggtcatcatgagctgcgcctg 336
QY 310 tgcctggtccgctgtggaaggaacattagaatgatgtgttccttaag 369
DB 337 tgcctggtccgctgtggaaggaacattagaatgatgtgttccttaag 396
QY 370 atggcagagaacagatcgtgtgtgattatttgaaggagattcaatttga 429
DB 397 atggcagagaacagatcgtgtgtgattatttgaaggagattcaatttga 456
QY 430 tgaagtcaacaagtgtgattcaatcattagaaggaacagcaggaatcttgatgc 489
DB 457 tgaagtca-caaagtgtgattcaatcattagaaggaacagcaggaatcttgatgc 515
QY 490 ttcaacaagatgcac 506
DB 516 ttcaacaagatgcac 532

RESULT 4
AAV62427 standard; cDNA; 2037 BP.
XX AAV62427;
AC

XX 30-DEC-1998 (first entry)
DT Prostate cancer antigen (PCA3) cDNA splice variant 1.
XX Prostate cancer antigen cDNA splice variant 1; PCA3; prostatic cancer;
DE PC; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
FH 379..534
FT GDS /*tag= a
FT /*product= "PCA3 protein variant 1"
FT 2019..2024
FT polyA_signal /*tag= b
XX W09845420-A1.
XX 15-OCT-1998.
XX 09-APR-1998; 98WO-CA00346.
XX 10-APR-1997; 97US-0041836.
XX (DIAG-) DIAGNOCURE INC.
XX Bussemakers MDG;
XX WPI, 1998-568347/48.
XX P-PSDB; AAV79736.
XX New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,
PT prevention and treatment of prostatic cancer
PT Claim 3; Fig 2B-2J; 111pp; English.
XX The present sequence represents the prostate cancer antigen (PCA3)
CC cDNA splice variant 1 sequence comprising of exons 1, 2, 3, 4a and
CC 4b of the PCA3 gene. The PCA3 cDNA splice variant 1 sequence,
CC isolated from a human primary prostatic tumor tissue cDNA library,
CC was found in approximately 5% of the cDNA clones isolated. The
CC invention claims for PCA3 cDNA variants and the proteins they encode.
CC The invention also claims for antibodies against PCA3 protein. The
CC antibodies are claimed to be useful for detecting PCA3 protein in
CC immunosay tests, for diagnosing, assessing and prognosing of
CC prostatic cancer (PC). Antibodies, optionally coupled to a cytotoxin
CC or radioisotope, and nucleic acids antisense to PCA3 cDNA are claimed
CC to be useful for treating PC, while determining elevated levels of
CC PCA3 (as RNA or protein) is useful for detecting a predisposition
CC to development of PC, e.g. in prenatal tests. Detecting PCA3 protein
CC allows differentiation between malignant and benign prostatic disease,
CC and the level of PCA3 expression allows correlation with the grade of
CC tumor. PCA3 protein and its fragments are also claimed to be useful
CC in vaccines for preventing PC; in drug screens for identifying
CC specific (antagonists (potentially useful therapeutically) and for
CC studying protein-DNA interactions.
XX Sequence 2037 BP; 622 A; 426 C; 406 G; 575 T; 8 other;
SQ

Query Match 48.1%; Score 243.4; DB 19; Length 2037;
Best Local Similarity 99.2%; Pred. No. 1.3e-69;
Matches 255; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 250 cacagatccctggagaatgcgcgcgcacatcttggtcatcatgagctgcgcctg 309
DB 442 cagagatccctggagaatgcgcgcgcacatcttggtcatcatgagctgcgcctg 501
QY 310 tgcctggtccgctgtggaaggaacattagaatgatgtgttccttaag 369
DB 502 tgcctggtccgctgtggaaggaacattagaatgatgtgttccttaag 561

```
OY 370 atggcaggaacacagatcctgtgtgatatatttgaacgggattacagatttgaa 429
    |||
DB 562 atggcaggaacacagatcctgtgtgatatatttgaacgggatttgaa 621
OY 430 tgaagtcacaaagtgagcatlaccacatgagaagaaacagagaaatcttgatg 489
    |||
DB 622 tgaagtcacaaagtgagcatlaccacatgagaagaaacagagaaatcttgatg 680
OY 490 ttcaacaagacatgcac 506
    |||
DB 681 ttcaacaagacatgcac 697

RESULT 5
AAA06688/c
ID AAA06688 standard; cDNA: 2229 BP.
XX
AC AAA06688;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:469.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW Immunogenic; cytosolic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PE 14-JUL-1999; 99MO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 260-261; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AA082000 to AA082020 represent sequences used in the exemplification of
CC the present invention.
XX
XX Sequence 2229 BP; 654 A; 447 C; 481 G; 647 T; 0 other;
XX
Query Match 48.1%; Score 243.4; DB 21; Length 2229;
Best Local Similarity 99.2%; Pred. NO. 1.4e-69;
Matches 255; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
```

```
OY 250 cacegacccctggagaaatgcccggccgacatcttggatcatgatgagcctgcctg 309
    |||
DB 1596 CAGAGATCCCTGGGAGAAATGCCGGCCGACATCTTGAGTCATGATGACCTGCGCCTG 1537
OY 310 tgctgtgtcccgcttgtgaggaaggaatagaatgaattatgttcccttaag 369
    |||
DB 1536 TGCCGTGTCGCCCTTGAGGGAAGGACATTAGAAAAGAAATGATGTGTCCTTAAAG 1477
OY 370 atggcaggaacacagatcctgtgtgatatatttgaacgggattacagatttgaa 429
    |||
DB 1476 ATGGCAGAGAAACAGATCCCTGTGTGATATTTATTGAAACGGGATTTACAGATTTGAA 1417
OY 430 tgaagtcacaaagtgagcatlaccacatgagaagaaacagagaaatcttgatg 489
    |||
DB 1416 TGAAGTCA-CAAGTGACATTACCAATGACAGAGAAAACAGACAGAAAATCTTGATGTC 1358
OY 490 ttcaacaagacatgcac 506
    |||
DB 1357 TTCACAAAGCATGCAAC 1341

RESULT 6
AAA06689/c
ID AAA06689 standard; cDNA: 2426 BP.
XX
AC AAA06689;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:470.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW Immunogenic; cytosolic; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PE 14-JUL-1999; 99MO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 1; Page 261-262; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
```


Oy	390	ttgttgatatttatttgaacggattacagatttgaatgaatgcacaaagtgagca	449
Db	174	tgctggatatttatttcttgaacggwtacacatttgaatgaagta-caaagtgaaca	232
Oy	450	ttaccacagagaggaacacagacggagaaatctgatgcttcacaaagatgcac	506
Db	233	ttaccacagagaggaacacagacggagaaatctgatgcttcacaaagatgcac	289
RESULT 10			
AC	AAA06545	AAA06545 standard; cDNA; 718 BP.	
XX	AAA06545;		
DT	13-JUN-2000	(first entry)	
XX			
DE	Human immunogenic prostate tumour protein cDNA sequence SEQ ID NO:313.		
KW	Human; prostate cancer; diagnosis; tumour; gene therapy; detection;		
KX	immunogenic; cytostatic; vaccine; ss.		
OS	Homo sapiens.		
PN	WO200004149-A2.		
PD	27-JAN-2000.		
PF	14-JUL-1999;	99WO-US15838.	
XX			
PR	14-JUL-1998;	98US-0115453.	
PR	14-JUL-1998;	98US-0116134.	
PR	23-SEP-1998;	98US-0159812.	
PR	23-SEP-1998;	98US-0159822.	
PR	15-JAN-1999;	99US-0232149.	
PR	15-JAN-1999;	99US-0232880.	
XX	09-APR-1999;	99US-0288946.	
PA	(CORI-) CORIXA CORP.		
PI	Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;		
DR	WPI; 2000-171268/15.		
XX			
PT	New polypeptide useful for treating and diagnosing prostate cancer		
PS	comprises an immunogenic portion of prostate tumor protein -		
XX			
PS	Claim 1; Page 199-200; 263pp; English.		
XX			
CC	The present invention describes isolated polypeptides, comprising an		
CC	immunogenic portion of a prostate tumour protein (PTP). The polypeptides		
CC	and polynucleotides encoding them have cytostatic activity and can be		
CC	used in vaccines and in gene therapy. The polypeptides and		
CC	polynucleotides encoding them, antigen presenting cells which express		
CC	the polypeptides, antibodies against the polypeptides and vaccines		
CC	comprising them can be used for inhibiting the development of prostate		
CC	cancer in a patient. The polypeptides can be used to generate antibodies		
CC	or anti-idiotypic antibodies for passive immuno therapy. A portion of		
CC	the polynucleotides encoding the polypeptides can be used as a probe or		
CC	to modulate the expression of the polypeptides. AAA06241 to AAA06691 and		
CC	AAA82000 to AAY82020 represent sequences used in the exemplification of		
CC	the present invention.		
XX			
SO	Sequence 718 BP; 222 A; 145 C; 169 G; 172 T; 10 other;		
Query Match 43.2%; Score 218.4; DB 21; Length 718;			
Best Local Similarity 98.1%; Pred. No. 1.2e-61;			
Matches 252; Conservative 0; Mismatches 2; Indels 3; Gaps 3;			
Oy	250	cacagatccctggagaaatgcgcggccacatttggatcatgatgagctgcctg	309
DB	11	ttgttgatatttatttgaacggattacagatttgaatgaatgcacaaagtgagca	449

Db	252	caggaatccctcggaggaatgccccggccgcatcttggjgcatcgaatgaagccctcgccctg	311
Qy	310	tgccctggtcccgcttgtagggaagagcatagaanaatgatgatgtgttctcttaag	369
Db	312	tgccctggtcccgcttgtagggaagagcatagaanaatgatgatgtgttctcttaag	371
Qy	370	atggcgcggagaaacagtcctgtgttgatattatttgaagcggtattgaatttgaa	429
Db	372	at-ggcagaaacagatccgctgtgtgataattatttgaagcggtattgaatttgaa	430
Qy	430	tgaagtcacccaagtgtgcatctacatgaatgaggaacacagacagagaanaatcttgatgc	489
Db	431	tgaagtcacccaagtgtgcatctacatgaatgaggaacacagacagagaanaatcttgatgc	488
Qy	490	ttcacagaacatgcgaac	506
Db	489	ttcacagaacatgcgaac	505
RESULT 11			
ID	AAc06768	AAc06768 standard; cDNA: 437 BP.	
xx	xx	AAc06768;	
xx	xx	06-OCT-2000 (first entry)	
xx	xx	Human secreted protein 5' EST, SEQ ID NO: 10843.	
xx	xx	Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;	
xx	xx	gene therapy; chromosome mapping; ss.	
xx	xx	Homo sapiens.	
xx	xx	EP1033401-A2.	
xx	xx	06-SEP-2000.	
xx	xx	21-FEB-2000; 2000EP-0200610.	
xx	xx	26-FEB-1999; 99US-0122487.	
xx	xx	(BEST) GENSET.	
xx	xx	Dumas Milne Edwards J, Duclert A, Giordano J;	
xx	xx	WPI: 2000-500381/45.	
xx	xx	New nucleic acid that is a 5' expressed sequence tag (5' EST) for	
xx	xx	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for	
xx	xx	diagnostic, forensic, gene therapy and chromosome mapping procedures -	
xx	xx	Claim 1; SEQ ID 10843; 71pp + CD-ROM; English.	
xx	xx	The present sequence is one of a large number of 5' ESTs derived from	
xx	xx	mRNAs encoding secreted proteins. No ORF has yet been conclusively	
xx	xx	identified within the present sequence. The 5' ESTs were prepared from	
xx	xx	total human RNAs or poly(A) RNAs derived from 30 different tissues. EST	
xx	xx	sequences usually correspond mainly to the 3' untranslated region (3'UTR)	
xx	xx	of the mRNA because they are often obtained from oligo-dT primed cDNA	
xx	xx	libraries. Such ESTs are not well suited for isolating cDNA sequences	
xx	xx	derived from the 5' ends of mRNAs and even in those cases where longer	
xx	xx	cDNA sequences have been obtained, the full 5' UTR is rarely included.	
xx	xx	5' ESTs are derived from mRNAs with intact 5' ends and can therefore be	
xx	xx	used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used	
xx	xx	in diagnostic, forensic, gene therapy and chromosome mapping procedures.	
xx	xx	They are used to obtain upstream regulatory sequences and to design	
xx	xx	expression and secretion vectors.	
xx	xx	Sequence 437 BP: 140 A; 87 C; 118 G; 88 T; 4 other;	
Query Match 27.3%; Score 138; DB 21; Length 437;			

Sequence 301 BP; 76 A; 58 C; 70 G; 97 T; 0 other;

Query Match	10.4%	Score 52.4	DB 20	length 359
Best Local Similarity	98.1%	Pred No. 2.5e-07		
Matches 53, Conservative	0	Mismatches 1	Indels 0	Gaps 0

250 cacagatcccttgaggagaatatgcccgcgcgcacatcttggtcatcatgagcctc 303

Db 306 cagagatccctggagaaatgcccgcgcacattgtggtcatgatgagctc 359

RESULT 14
AAF59232/c
ID AAF59232 standard; DNA; 1515 BP.
AC AAF59232;
XX
XX 26-APR-2001 (first entry)
DE Human PTH2 receptor ligand TIP39 related genomic DNA sequence.
XX
XX Tuberin/fundibular peptide of 39 residues; TIP39; PTH2 receptor ligand;
KW parathyroid hormone; parathyroid hormone receptor; anorectic; nootropic;
KW PTH2 receptor binding activity; antidepressant; neuroleptic; analgesic;
KW antimigraine; antidiabetic; osteopathic; hypertensive; cardiac;
KW cytosolic; antistatic; neuroprotective; PTH receptor antagonist;
KW obesity; eating disorder; metabolic disorder; mental disorder;
KW depression; schizophrenia; dementia; acute pain; chronic pain; migraine;
KW headache; diabetes; osteoporosis; hypercalcaemia; calcium metabolism;
KW hyperextension; congestive heart failure; tumour; asthma; emphysema;
KW restrictive lung disease; demyelinating condition; multiple sclerosis;
KW leukodystrophy; ds.
XX
XX Homo sapiens.
OS
XX WO200077042-A2.
PN
XX 21-DEC-2000.
PD
XX 15-JUN-2000; 2000WO-US16776.
PF
XX 15-JUN-1999; 99US-0139335.
PR
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA
XX Usdin TB, Hoare SRJ;
PI
XX WPI; 2001-122833/13.
DR
XX
XX New parathyroid hormone type 2 or 1 receptor ligand, useful for
PT treating e.g. migraine or headaches, hypertension, obesity and other
PT eating or metabolic disorders, mental disorders and osteoporosis -
XX
XX Example 9; Page 78-79; 106pp; English.

The present invention describes an isolated or purified peptide (I) that is a parathyroid hormone type 2 (PTH2) receptor ligand (Ia) or PTH1 receptor ligand (Ib). (I) has anorectic, neuroleptic, antidepressant, nootropic, analgesic, antimigraine, antidiabetic, osteopathic, hypertensive, cardiac, cytosolic, antistatic, antistatic and neuroprotective activities, and is an PTH receptor antagonist. The peptide is useful in the treatment of a disorder mediated by binding a PTH2 or PTH1 receptor. The PTH2 receptor binding activity may be used in treating obesity or other eating or metabolic disorders, mental disorders (e.g. depression, schizophrenia and dementia), acute or chronic pain, migraine or headaches, diabetes and other metabolic disorders, osteoporosis, hypercalcaemia and other disorders affecting calcium metabolism, hypertension, congestive heart failure and control of tumour growth, asthma, emphysema or other restrictive lung diseases, and demyelinating conditions such as multiple sclerosis and leukodystrophies. The present sequence represents a PTH2 receptor ligand TIP39 related genomic DNA sequence, which is given in the exemplification of the present invention.

Sequence 1515 BP; 269 A; 451 C; 518 G; 277 T; 0 other;

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 16, 2001, 17:57:37 ; Search time 51.36 Seconds
(without alignments)
27.149 Million cell updates/sec

Title: us-09-675-650-3
Perfect score: 123
Sequence: 1 MFHISPFKYPHTQEAQKRAQR 23

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.GeneSeq.0601.*
1: /SIDSB/gcgdata/geneSeq/geneSeq/AA1980.DAT.*
2: /SIDSB/gcgdata/geneSeq/geneSeq/AA1981.DAT.*
3: /SIDSB/gcgdata/geneSeq/geneSeq/AA1982.DAT.*
4: /SIDSB/gcgdata/geneSeq/geneSeq/AA1983.DAT.*
5: /SIDSB/gcgdata/geneSeq/geneSeq/AA1984.DAT.*
6: /SIDSB/gcgdata/geneSeq/geneSeq/AA1985.DAT.*
7: /SIDSB/gcgdata/geneSeq/geneSeq/AA1986.DAT.*
8: /SIDSB/gcgdata/geneSeq/geneSeq/AA1987.DAT.*
9: /SIDSB/gcgdata/geneSeq/geneSeq/AA1988.DAT.*
10: /SIDSB/gcgdata/geneSeq/geneSeq/AA1989.DAT.*
11: /SIDSB/gcgdata/geneSeq/geneSeq/AA1990.DAT.*
12: /SIDSB/gcgdata/geneSeq/geneSeq/AA1991.DAT.*
13: /SIDSB/gcgdata/geneSeq/geneSeq/AA1992.DAT.*
14: /SIDSB/gcgdata/geneSeq/geneSeq/AA1993.DAT.*
15: /SIDSB/gcgdata/geneSeq/geneSeq/AA1994.DAT.*
16: /SIDSB/gcgdata/geneSeq/geneSeq/AA1995.DAT.*
17: /SIDSB/gcgdata/geneSeq/geneSeq/AA1996.DAT.*
18: /SIDSB/gcgdata/geneSeq/geneSeq/AA1997.DAT.*
19: /SIDSB/gcgdata/geneSeq/geneSeq/AA1998.DAT.*
20: /SIDSB/gcgdata/geneSeq/geneSeq/AA1999.DAT.*
21: /SIDSB/gcgdata/geneSeq/geneSeq/AA2000.DAT.*
22: /SIDSB/gcgdata/geneSeq/geneSeq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123	100.0	51	19	AAW79736
2	123	100.0	51	19	AAW79738
3	56	45.5	11	19	AAW79737
4	46.5	37.8	11	19	AAW96038
5	44.5	36.2	234	22	AAW20288
6	44.5	36.2	419	21	AAW08021
7	44.5	36.2	511	17	AAW04539
8	44.5	36.2	511	20	AAW23748
9	44.5	36.2	511	20	AAW73506
10	44.5	36.2	511	22	AAW20290
11	44.5	36.2	511	22	AAW59298

12	44	35.8	492	19	AAW56281	Babesia microti BM
13	44	35.8	492	20	AAW24342	Babesia microti an
14	44	35.8	492	21	AAW30191	B. microti BMN1-4
15	44	35.8	503	19	AAW56302	Babesia microti MN
16	44	35.8	503	20	AAW24364	Babesia microti an
17	44	35.8	503	21	AAW30216	B. microti clone a
18	43.5	35.4	166	21	AAW5637	Human colon cancer
19	43.5	35.4	245	8	AAW70677	Human kallikrein g
20	43.5	35.4	245	10	AAW92314	Human recombinant k
21	43.5	35.4	256	21	AAW21315	Human KLK1. Homo
22	43.5	35.4	262	8	AAW70568	Human kallikrein-1
23	43.5	35.4	262	10	AAW95121	Kallikrein encoded
24	43.5	35.4	262	19	AAW71005	Human prostate-ess
25	43.5	35.4	262	21	AAW21319	Human KLK2. Homo
26	43.5	35.4	267	21	AAW54293	Human pancreatic c
27	43.5	35.4	511	22	AAW20289	Vesicular stomatit
28	43.5	35.0	177	21	AAW85637	Monkey fibroblast
29	43	35.0	178	21	AAW85638	Human fibroblast g
30	43	35.0	178	21	AAW85639	Human fibroblast g
31	43	35.0	250	21	AAW43790	Arabidopsis thalia
32	43	35.0	251	22	AAW65297	Arabidopsis thalia
33	43	35.0	363	21	AAW77125	Human neurotransm
34	43	35.0	585	20	AAW23643	Protein encoded by
35	42	34.1	127	20	AAW24088	Salmonella typhimu
36	42	34.1	188	21	AAW58237	Arabidopsis thalia
37	42	34.1	232	21	AAW58236	Arabidopsis thalia
38	42	34.1	247	21	AAW58235	Arabidopsis thalia
39	41.5	33.7	364	18	AAW3643	Thermotabile alpha
40	41.5	33.7	431	12	AAW17331	Human cyclin A. H
41	41.5	33.7	515	21	AAW56707	Human prostate can
42	41	33.3	119	11	AAW09304	Sequence deduced f
43	41	33.3	167	21	AAW22828	Arabidopsis thalia
44	41	33.3	178	21	AAW41814	Human ORF1578
45	41	33.3	250	21	AAW94965	Human secreted pro

ALIGNMENTS

RESULT 1	AAW79736	standard; Protein; 51 AA.
ID	AAW79736	
AC	AAW79736	
XX		
XX		
DT	30-DEC-1998	(first entry)
XX		
DE	Prostate cancer antigen (PCA3) protein variant 1.	
KW	Prostate cancer antigen protein variant 1; PCA3; prostatic cancer;	
KW	PC.	
OS	Homo sapiens.	
PN	W09845420-A1.	
PD	15-OCT-1998.	
PF	09-APR-1998;	98WO-CA00346.
PR	10-APR-1997;	97US-0041836.
PA	(DIAG-) DIAGNOCURE INC.	
PI	Bussemakers MGJ;	
XX		
XX		
DR	MP1: 1998-568347/48.	
XX	N-PSDB: AAW62427.	
XX	New nucleic acid encoding prostate cancer antigen 3 - for diagnosis,	
XX	prevention and treatment of prostatic cancer	
PS	Claim 16; Fig 2B-2J; 111pp; English.	

45 108 #1 cultured

in vaccines for preventing PC; in drug screens for identifying specific (ant)agonists (potentially useful therapeutically) and for studying protein-DNA interactions.

Sequence 11 AA:

Query Match 45.5%; Score 56; DB 19; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0096;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 HTOEAKPEAOR 23
| | | | | | | | | | | | | | |
Db 1 htgeaqkeagr 11

RESULT 4

AAR96038
ID AAR96038 standard; Protein: 719 AA.

AC AAR96038;

DT 24-FEB-1997 (first entry)

DE 80 kD protein.

Tetrahymena; 80 kD protein; 95 kD protein; telomerase; telomeric repeat; RNA coupled protein; chromosome; telomere repeat synthesis; antibody; immortalised cell; cancer; eukaryotic microbe; inhibitor; gene therapy; fungal infection; therapy; diagnosis; protozoan infection; somatic cell; p80; p95.

Tetrahymena thermophila.

WO9619580-A2.

27-JUN-1996.

18-DEC-1995; 95WO-US16531.

19-DEC-1994; 94US-0359125.

(COLD-) COLD SPRING HARBOR LAB.

Autexier C, Collins K, Greider C, Hemish JM, Kobayashi R;

Yang XH;

WPI: 1996-309594/31.

N-PSDB: AAT30867.

Telomerase protein and related DNA, antibodies, transgenic cells, etc. - for diagnosis and treatment of cancer and infection by eukaryotic microbes, also new telomerase inhibitors

Claim 10; Fig 2; 56pp; English.

AAR96038 and AAR96039 represent Tetrahymena 80 kD and 95 kD proteins (p80 and p95), respectively. These proteins are components of telomerase. The proteins act, when coupled to RNA, to add telomeric repeats (of the sequence TTGGGG) to stabilise chromosomes. The RNA component provides the template for the telomere repeat synthesis. Antibodies against these sequences, can be used as immunoassay reagents for the detection of the proteins. The antibodies can also be used to identify immortalised cells, or predispotion to immortalisation, particularly cancer, or to diagnose a disease caused by a eukaryotic microbe. Inhibitors of these proteins (and these proteins themselves) can be used for therapy or diagnosis. As somatic cells do not generally require telomerase, these inhibitors should have little or no toxicity to the host. The DNA encoding these sequences are used to produce the recombinant protein, or to isolate similar genes from other organisms, while transformed cells can be used in gene therapy.

Sequence 719 AA:

Query Match 37.8%; Score 46.5; DB 17; Length 719;
Best Local Similarity 41.4%; Pred. No. 34;
Matches 12; Conservative 1; Mismatches 7; Indels 9; Gaps 1;

OY 2 FLHISP-----FKYPHTGEAKKA 21
| | | | | | | | | | | | | | |
Db 298 Ichisepkerykllgkypkreeykaa 326

RESULT 5

AAB20288
ID AAB20288 standard; Protein: 234 AA.

AC AAB20288;

DT 29-MAY-2001 (first entry)

DE Vesicular stomatitis virus mutant M2 G protein (partial).

VSV; G protein; tumour; ovarian carcinoma; lung carcinoma; colon carcinoma; leukaemia; lymphoma; myeloma; glioblastoma; melanoma; sarcoma; neuroendocrine tumour; cancer; therapy; oncolytic; cytostatic; antitumour; mutant; mutein.

Vesicular stomatitis virus.

WO200119380-A2.

22-MAR-2001.

18-SEP-2000; 2000WO-US25510.

17-SEP-1999; 99US-0397873.

(PROV-) PRO-VIRUS INC.

Bell JC, Sonenberg N, Stojdl DF, Brown EG, Atkins HL, Marius RM;

Lichty BD, Knowles SB;

WPI: 2001-244699/25.

N-PSDB: AAF30425.

Reducing the viability of carcinoma, melanoma, or sarcoma cells comprises administering a vesicular stomatitis virus able to selectively infect and kill tumour cells in a population of tumour and non-tumour cells -

Example 27; Fig 21/2-21/2; 118pp; English.

This is a partial sequence (missing N-terminal region) of the G protein of vesicular stomatitis virus (VSV) Indiana strain mutant M2. The invention is directed to a method of reducing the viability of a tumour cell by administering a virus, such as VSV, that is not a common human pathogen. VSV has a broad host range and is capable of infecting most types of human cells. It is an RNA virus that spends its entire life cycle in the cytoplasm. Therefore, it involves less danger of unwanted integration into the genome of a patient. The ability of VSV to selectively infect tumour cells over wild-type cells was observed. VSV mutants were isolated, including M2, that grew poorly on interferon responsive cells. The rationale was that VSV mutants, which can induce interferon in target cells, would limit their own replication in an interferon responsive cell population. These same viruses would however have unrestricted growth in tumour cells that lack interferon responsiveness. These mutants have even less cytopathic effect on normal tissues while maintaining the oncolytic activity of the wild-type VSV. They kill tumour cells efficiently while sparing normal cells, and also have the ability to produce more virus particles and increase virus spread throughout the tumour. Suitable tumours for treatment are ovarian carcinoma, lung carcinoma,

CC colon carcinoma, hematopoietic tumour, glioblastoma, melanoma,
 CC sarcoma, fibrosarcoma, neuroendocrine tumour, acute myelogenous
 CC leukaemia, chronic myelogenous leukaemia, promyelocytic leukaemia,
 CC lymphoma and melanoma (all claimed).

SO Sequence 234 AA;

Query Match 36.2%; Score 44.5; DB 22; Length 234;
 Best Local Similarity 42.9%; Pred. No. 20;
 Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

OY 3 LHISSP---FKYPHTQEAQKE 20
 Db 135 lhlsskagvfephqdaasg 155

RESULT 6

AAB08021 standard; Protein: 419 AA.

AC AAB08021;

DT 14-NOV-2000 (first entry)

DE A microphthalmia-associated transcription factor (Mtf).

KW Microphthalmia-associated transcription factor; Mtf; Mtf(+); Mtf(-);
 splice variant; cancer cell; cancer cell marker; melanoma.

OS Homo sapiens.

Key Location/Qualifiers

FT Misc-difference 187..192
 FT "these amino acids are uniquely present in
 splice variant Mtf(+)"

WT0200047765-A1.

PD 17-AUG-2000.

PE 03-FEB-2000; 2000MO-GB00313.

PR 08-FEB-1999; 99GB-0002776.

PA (CURIE-) CURIE CANCER CARE MARIE.

PI Coding CR;

DR WPI: 2000-549153/50.

DR N-PSDB; AAA59714.

PT Novel method for the detection of melanoma comprises determining the
 PT predominance of the splice variants Mtf (-) and Mtf (+) in a
 PT biological sample -

PS Disclosure; Fig 1; 29pp; English.

CC The present sequence represents a microphthalmia-associated transcription
 CC factor (Mtf). The protein exists in 2 forms, Mtf(+) and Mtf(-).
 CC Amino acids 187-192 are uniquely present in the Mtf(+) splice variant.
 CC In normal cells, Mtf(+) RNA is predominantly present, while in certain
 CC cancer cell lines, Mtf(-) is predominantly present. The specification
 CC describes a method for determining the predominance of either of the
 CC splice variants Mtf(-) and Mtf(+) in a biological sample. The method
 CC comprises independently detecting Mtf(-) and Mtf(+), and then
 CC comparing the amounts of the spliced variants in order to determine
 CC which variant is predominant in the sample. The method is used to detect
 CC the presence of and determine the predominance of the two splice variants
 CC of the cancer cell marker Mtf. Determination of these parameters leads
 CC to an indication of the presence, or likelihood of developing cancer,
 CC especially a melanoma.

SO Sequence 419 AA;

Query Match 36.2%; Score 44.5; DB 21; Length 419;
 Best Local Similarity 47.1%; Pred. No. 39;
 Matches 8; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

OY 4 HISSPFKYPHTQEAQKE 20
 Db 15 hlephtky-hlqqaqrq 30

RESULT 7

AAW04539 standard; Protein: 511 AA.

AC AAW04539;

DT 26-JAN-1997 (first entry)

DE Vesiculovirus glycoprotein (G).

KW Rhabdovirus; VSV; vaccine; PVSVFL(+); vector; immunotherapy;
 antigen; tumour; vesiculovirus.

OS Vesicular stomatitis virus.

Key Location/Qualifiers

FT Domain 456..482
 FT /label= "Transmembrane_domain
 483..511
 /label= Cytoplasmic_domain

FT Domain

WT09634625-A1.

PD 07-NOV-1996.

PE 01-MAY-1996; 96MO-US06053.

PR 04-MAY-1995; 95US-0435032.

PA (UYVA) UNIV YALE.

PI Rose JK;

DR WPI: 1996-505901/50.

DR N-PSDB; AAT38744.

PT Recombinant replicable vesiculovirus - useful to develop prods. for
 PT diagnosis and therapy, partic. in vaccines for infections, diseases
 PT or tumours

PS Disclosure; Page 91-92; 162pp; English.

CC A single glycoprotein (G) species (AAW04540) spans the membrane of
 CC vesiculovirus (VSV) and forms spikes on the surface of the virus
 CC particle. 3 Proteins, termed N or nucleocapsid (AAW04536), P
 CC (formerly termed NS indicating non-structural) (AAW04537) and L or
 CC large (AAW04540) are associated with the VSV nucleocapsid. An
 CC additional matrix (M) protein (AAW04538) lies within the membrane
 CC envelope. The proteins are encoded by VSV(-) DNA in PVSVFL(+)
 CC (AAT38744), a plasmid that can be used in the prodn. of recombinant,
 CC replicable VSV for use as vaccines.

SO Sequence 511 AA;

Query Match 36.2%; Score 44.5; DB 17; Length 511;
 Best Local Similarity 42.9%; Pred. No. 48;
 Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

OY 3 LHISSP---FKYPHTQEAQKE 20
 lhlsskagvfephqdaasg 155

Db 412 Jhlskagvfephidqdaasq 432

RESULT 8

AAW23748 standard; Protein; 511 AA.

AAW23748;

08-SEP-1999 (first entry)

Vesicular stomatitis virus G stem polypeptide.

Rhabdovirus; heterologous fusion protein; virus fusion; target cell membrane; N protein; P protein; L protein; G stem polypeptide; anti-receptor protein; infection; human immune deficiency virus; neoplastic condition; benign tumour; polyp; cafe au lait spot; leukoplakia; skin mole.

Vesicular stomatitis virus.

WO9932648-A1.

01-JUL-1999.

22-DEC-1998; 98WO-US26084.

22-DEC-1997; 97US-0068472.

(UYTE-) UNIV TENNESSEE RES CORP.

Robinson CS, Whitt MA;

WPI; 1999-418938/35.

Recombinant rhabdovirus containing heterologous fusion protein

Disclosure; Page 14; 63pp; English.

The specification describes genetically engineered rhabdoviruses which comprise a heterologous F (fusion) protein, or its fragment, to facilitate fusion of the virus with a target cell membrane, and the rhabdovirus N, P and L proteins. Recombinant viruses which include a G stem polypeptide and an anti-receptor protein (ARP) are used to fuse with target cells (depending on nature of ARP), for treatment of diseases associated with a surface protein recognized by ARP. Particularly the method is used to treat viral, parasitic and bacterial infections (specifically human immune deficiency virus); (pre-)neoplastic conditions; benign tumours; polyps; cafe au lait spots; leukoplakia and skin moles. Cells may be treated in vivo or in vitro. Recombinant rhabdoviruses that infect diseased or abnormal cells and also express a reporter can be used for diagnosis, detection and monitoring progression/regression of disease, also for studying function and specificity of proteins not naturally found on rhabdovirus. The present sequence may be used to create the rhabdoviruses of the invention.

Sequence 511 AA;

Query Match 36.2%; Score 44.5; DB 20; Length 511;

Best Local Similarity 42.9%; Pred. No. 48; Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 3 LHISSP---FKYPHTQEAQKE 20

Db 412 Jhlskagvfephidqdaasq 432

RESULT 9
AAW73506
ID AAW73506 standard; Protein; 511 AA.
XX AAW73506;
AC

XX 01-MAR-1999 (first entry)

DT VSV-G protein sequence.

DE DAF; decay accelerating factor; human; complement protein; gene therapy;

KW viral vector; VSV-G protein.

XX Vesicular stomatitis virus.

XX JP10313865-A.

XX 02-DEC-1998.

XX 15-MAY-1997; 97JP-0125965.

XX 15-MAY-1997; 97JP-0125965.

XX (DINA-) DINABEKKU KENKUYUSHO KK.

XX WPI; 1999-074147/07.

XX N-PSDB; AAV08936.

XX Vector having complement controlling factor - useful for gene therapy

XX Example 3; Page 9-11; 15pp; Japanese.

XX This sequence represents the VSV-G protein. DNA encoding this sequence can be used in the viral vector of the invention. The viral vector contains a factor controlling the function of a human complement protein, particularly a membrane combining type protein. The viral vector, which is stable in situ, is useful for gene therapy.

XX Sequence 511 AA;

SQ

Query Match 36.2%; Score 44.5; DB 20; Length 511;

Best Local Similarity 42.9%; Pred. No. 48; Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 3 LHISSP---FKYPHTQEAQKE 20

Db 412 Jhlskagvfephidqdaasq 432

RESULT 10

AAB20290 standard; Protein; 511 AA.

AAB20290;

29-MAY-2001 (first entry)

Vesicular stomatitis virus mutant M4 G protein.

VSV; G protein; tumour; ovarian carcinoma; lung carcinoma; colon carcinoma; leukaemia; lymphoma; glioblastoma; melanoma; sarcoma; neuroendocrine tumour; cancer; therapy;

KW oncolytic; cytostatic; antitumour; mutant; mutein.

XX Vesicular stomatitis virus.

XX

XX Key Location/Qualifiers

XX Misc-difference 294..348 /label= Unknown

XX Misc-difference 254 /note= "encoded by incompletely sequenced portion of N gene"

XX MO200119380-A2.

PD 22-MAR-2001.
XX
PF 18-SEP-2000; 2000WO-US25510.
XX
PR 17-SEP-1999; 99US-0397873.
XX
PA (PROV-) PRO-VIRUS INC.
PI Bell JC, Sorenberg N, Stojdl DF, Brown EG, Atkins HL, Marius RM;
PI Llochy BD, Knowles SB;
XX
XX WPI: 2001-244699/25.
DR N-PSDB; AAF30427.
XX
XX Reducing the viability of carcinoma, melanoma, or sarcoma cells
PT comprises administering a vesicular stomatitis virus able to
PT selectively infect and kill tumour cells in a population of tumour and
PT non-tumour cells -
PS
PS Example 27: Fig 21/1-21/2; 118pp; English.
XX
XX The present sequence is that of the G protein from vesicular
CC stomatitis virus (VSV) Indiana strain mutant M4. This mutant
CC sequence has a single amino acid change, Gly234 rather than Glu,
CC compared with that of a wild-type heat resistant strain G protein
CC (see AAB20287). Such changes in sequence correlate with phenotypic
CC differences of the VSV. The invention is directed to a method of
CC reducing the viability of a tumour cell by administering a virus,
CC such as VSV, that is not a common human pathogen. VSV has a broad
CC host range and is capable of infecting most types of human cells. It
CC is an RNA virus that spends its entire life cycle in the cytoplasm.
CC Therefore, it involves less danger of unwanted integration into the
CC genome of a patient. The ability of VSV to selectively infect
CC tumour cells over wild-type cells was observed. VSV mutants were
CC isolated, including M4, that grew poorly on interferon responsive
CC cells. The rationale was that VSV mutants, which can induce
CC interferon in target cells, would limit their own replication in an
CC interferon responsive cell population. These same viruses would
CC however have unrestricted growth in tumour cells that lack interferon
CC responsiveness. These mutants have even less cytopathic effect on
CC normal tissues while maintaining the oncolytic activity of the
CC wild-type VSV. They kill tumour cells efficiently while sparing
CC normal cells, and also have the ability to produce more virus
CC particles and increase virus spread throughout the tumour. Suitable
CC tumours for treatment are ovarian carcinoma, lung carcinoma,
CC colon carcinoma, haematopoietic tumour, glioblastoma, melanoma,
CC sarcoma, fibrosarcoma, neuroendocrine tumour, acute myelogenous
CC leukaemia, chronic myelogenous leukaemia, promyelocytic leukaemia,
CC lymphoma and melanoma (all claimed).
XX
XX Sequence 511 AA;
SQ

Query Match 36.2%; Score 44.5; DB 22; Length 511;
Best Local Similarity 42.9%; Pred. No. 48;
Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;
QY 3 LHISP---FKYPHQEAKKE 20
||:| |::| |::| |:
DB 412 JHLSKQYFHPHQDAQS 432

RESULT 11
AAB59298
ID AAB59298 standard; protein: 511 AA.
XX
XX AAB59298;
AC
XX
XX 02-APR-2001 (first entry)
DT
XX
XX Vesicular stomatitis virus G protein.
DE
XX
XX Vesiculovirus; vaccine; infection; cancer.
KW

XX
OS Vesicular stomatitis virus.
XX
XX US6168943-B1.
XX
XX 02-JAN-2001.
PD
XX
XX 03-MAY-1996; 96US-0646695.
PF
XX
XX 04-MAY-1995; 95US-0435032.
PR
XX
XX (UYRA) UNITV YALE.
PA
XX
XX Rose JK;
PI
XX
XX WPI: 2001-136716/14.
DR
XX
XX
XX
XX Producing recombinant replicable vesiculovirus, useful as vaccines for
PT treating or preventing microbial infections, comprises culturing a cell
PT containing a nucleic acid for the expression of vesiculovirus
PT antigenomic RNA -
PS
PS Disclosure: Fig 1; 119pp; English.
XX
XX The present invention relates to producing a recombinant replicable
CC vesiculovirus. The method involves culturing a cell containing a first
CC recombinant nucleic acid that can be transcribed to produce an RNA
CC comprising a vesiculovirus antigenomic (+) RNA containing the
CC vesiculovirus promoter for replication and a ribozyme sequence
CC immediately downstream the antigenomic (+) RNA. The method is useful
CC for producing recombinant replicable vesiculoviruses, which can
CC be used as vaccines for the treatment or prevention of
CC infections by a pathogenic microorganism. The recombinant
CC replicable vesiculoviruses are useful in diagnosing and monitoring
CC progression of infectious disorders, including response to vaccination
CC and/or therapy, in cancer immunoprophylaxis, immunotherapy and to
CC diagnosis, and monitoring of tumor progression or regression, and to
CC produce large quantities of readily purified antigen.
XX
XX Sequence 511 AA;
SQ

Query Match 36.2%; Score 44.5; DB 22; Length 511;
Best Local Similarity 42.9%; Pred. No. 48;
Matches 9; Conservative 5; Mismatches 4; Indels 3; Gaps 1;
QY 3 LHISP---FKYPHQEAKKE 20
||:| |::| |::| |:
DB 412 JHLSKQYFHPHQDAQS 432

RESULT 12
AAM56281
ID AAM56281 standard; protein: 492 AA.
XX
XX AAM56281;
AC
XX
XX 28-SEP-1998 (first entry)
DT
XX
XX Babesia microti BMNT-4 antigen sequence.
DE
XX
XX antigen; detection; diagnosis; vaccine; tick-borne disease;
KW differentiation; Lyme disease; ehrlichiosis.
XX
XX Babesia microti.
OS
XX
XX EP834567-A2.
FN
XX
XX 08-APR-1998.
PD
XX
XX 01-OCT-1997; 97EP-0117067.
PF
XX
XX 24-APR-1997; 97US-0845256.
PR

PR 01-OCT-1996; 96US-0723142.
 XX
 XX (CORI-) CORIXA CORP.
 XX
 PI Houghton R, Lodes MJ, Reed SG, Sleath PR;
 XX
 DR WPI: 1998-195465/18.
 N-PSDB; AAY22734.
 XX
 XX Polypeptides comprising Babesia microti antigens and their
 PT immunogenic fragments or epitopes - and related nucleic acid,
 PT vectors, transformed cells and antibodies, useful for diagnosis of
 PT infection and in protective vaccines
 XX
 PS Claim 1; Page 49-51; 113pp; English.
 XX
 XX The sequence is that of a polypeptide comprising at least
 CC one antigenic portion of a Babesia microti antigen. It can be used
 CC to diagnose B. microti infection by detecting specific antibodies
 CC in usual immunoassays. Infection can also be diagnosed using:
 CC (a) primers or probes derived from the coding sequence, in
 CC standard amplification or hybridisation tests, or (b) using
 CC antibodies to detect the corresponding antigen. It is also
 CC useful in vaccines to protect against infection, especially
 CC when formulated with an adjuvant. The new diagnostic methods
 CC allow rapid differentiation between B. microti infection and
 CC other tick-borne diseases (Lyme disease and ehrlichiosis) that
 CC have similar symptoms but require different treatments.
 CC
 SQ Sequence 492 AA;

Query Match 35.8%; Score 44; DB 19; Length 492;
 Best Local Similarity 31.8%; Pred. No. 56;
 Matches 7; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 FLHISPFKYPHTEOAKQKOR 23
 | : | | : : | | : : | : : :
 Db 103 fryiskeyehetelakehck 124

RESULT 13
 AAY24342
 ID AAY24342 standard; Protein; 492 AA.
 XX
 AC AAY24342;
 XX
 DT 16-SEP-1999 (first entry)
 XX
 DE Babesia microti antigen BMN1-4.
 XX
 KW Babesia microti; antigen; immunogen; diagnosis; infection; vaccine;
 XX immunity; detection.
 XX
 OS Babesia microti.
 XX
 PN WO9929869-A1.
 XX
 PD 17-JUN-1999.
 XX
 PF 11-DEC-1998; 98MO-US26437.
 XX
 PR 11-DEC-1997; 97US-0990571.
 XX
 PA (CORI-) CORIXA CORP.
 PA (MAYO-) MAYO FOUNDATION.
 XX
 PI Brunsmma E, Houghton R, Lodes MJ, Persing D, Reed SG;
 PI Sleath PR;
 XX
 XX WPI: 1999-385612/32.
 DR N-PSDB; AAX88998.
 XX

PT New isolated Babesia microti polypeptides
 XX
 PS Example 1; Page 72-74; 126pp; English.
 XX
 CC The present invention describes isolated polypeptides comprising
 CC specific immunogenic portions of Babesia microti. AAX88983 to AAX88994
 CC encode specifically claimed B. microti immunogenic proteins, and
 CC AAY24327 to AAY24338 represent these proteins. B. microti polypeptides
 CC and nucleic acids can be used for detecting B. microti infections. They
 CC can also be used in vaccines for inducing protective immunity against B.
 CC microti infections. The present sequence represents a B. microti antigen.
 XX
 SQ Sequence 492 AA;

Query Match 35.8%; Score 44; DB 20; Length 492;
 Best Local Similarity 31.8%; Pred. No. 56;
 Matches 7; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 FLHISPFKYPHTEOAKQKOR 23
 | : | | : : | | : : | : : :
 Db 103 fryiskeyehetelakehck 124

RESULT 14
 AAB30191
 ID AAB30191 standard; Protein; 492 AA.
 XX
 AC AAB30191;
 XX
 DT 12-FEB-2001 (first entry)
 XX
 DE B. microti BMN1-4 antigen SEQ ID NO: 21.
 XX
 KW Babesiosis; rodent parasite; tick-borne illness; antigen;
 XX disease diagnosis; disease prevention.
 XX
 OS Babesia microti.
 XX
 PN WO200060090-A1.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000MO-US09136.
 XX
 PR 05-APR-1999; 99US-0286488.
 XX
 PR 17-MAR-2000; 2000US-0528784.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD;
 PI WPI: 2000-686939/67.
 DR N-PSDB; AAC65081.
 XX
 XX New polypeptides containing an antigenic portion of Babesia microti
 PT antigen and DNAs encoding the polypeptides, useful for diagnosing,
 PT treating or preventing B. microti infection, or for inducing protective
 PT immunity in a patient
 XX
 PS Disclosure; Page 72-73; 118pp; English.
 XX
 CC The present invention is related to the isolation of antigenic sequences
 CC from the rodent parasite Babesia microti. This organism is transmitted to
 CC humans by the same tick which transmits Lyme disease and ehrlichiosis.
 CC The organism causes a malaria-like infection known as babesiosis. The
 CC sequences identified by this invention can be used in the diagnosis,
 CC prevention and treatment of babesiosis.
 XX
 SQ Sequence 492 AA;

Query Match 35.8%; Score 44; DB 21; Length 492;

Best Local Similarity 31.8%; Pred. No. 56;
Matches 7; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Job time: 3896 sec

OY 2 FLHISPPKYPHTQEAQKEAQR 23
| : | | : | | : | : : :
Db 103 frylskeyeyentelakehckk 124

RESULT 15

AAW56302
ID AAW56302 standard; Protein; 503 AA.

XX AAW56302;

XX 28-SEP-1998 (first entry)

XX Babesia microti MN-10 antigen sequence.

XX antigen; detection; diagnosis; vaccine; tick-borne disease;

KW differentiation; Lyme disease; ehrlichiosis.

XX Babesia microti.

XX EP834567-A2.

XX 08-APR-1998.

XX 01-OCT-1997; 97EP-0117067.

XX 24-APR-1997; 97US-0845258.

XX 01-OCT-1996; 96US-0723142.

XX (CORI-) CORIXA CORP.

XX Houghton R, Lodes MJ, Reed SG, Sleath PR;

XX WPI; 1998-195465/18.

XX N-PSDB; AAV22748.

PT Polypeptides comprising Babesia microti antigens and their
PT immunogenic fragments or epitopes - and related nucleic acid,
PT vectors, transformed cells and antibodies, useful for diagnosis of
PT infection and in protective vaccines

PS Claim 1; Page 99-101; 113pp; English.

CC The sequence is that of a polypeptide comprising at least
CC one antigenic portion of a Babesia microti antigen. It can be used
CC to diagnose B. microti infection by detecting specific antibodies
CC in usual immunoassays. Infection can also be diagnosed using:
CC (a) primers or probes derived from the coding sequence, in
CC standard amplification or hybridisation tests, or (b) using
CC antibodies to detect the corresponding antigen. It is also
CC useful in vaccines to protect against infection, especially
CC when formulated with an adjuvant. The new diagnostic methods
CC allow rapid differentiation between B. microti infection and
CC other tick-borne diseases (Lyme disease and ehrlichiosis) that
CC have similar symptoms but require different treatments.

XX Sequence 503 AA:

Query Match 35.8%; Score 44; DB 19; Length 503;

Best Local Similarity 31.8%; Pred. No. 57;

Matches 7; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

OY 2 FLHISPPKYPHTQEAQKEAQR 23
| : | | : | | : | : : :
Db 31 frylskeyeyentelakehckk 52

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 17:51:07 ; Search time 1566.45 Seconds
(without alignments)
4996.445 Million cell updates/sec

Title: US-09-675-650-1
Perfect score: 506
Sequence: 1 caggagacacacaaaggagac.....ggcttcacacagacatgcaac 506

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 segs, 773874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_cm:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pi1:*
- 13: gb_pi2:*
- 14: gb_pi3:*
- 15: gb_pi4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_cm:*
- 43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pi:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_v1:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_v11:*

59: gb_v12:*

60: gb_hc91:*

61: gb_hc92:*

62: gb_hc93:*

63: gb_hc94:*

64: gb_hc95:*

65: gb_hc96:*

66: gb_hc97:*

67: gb_hc98:*

68: gb_hc99:*

69: gb_hc10:*

70: gb_hc911:*

71: gb_hc912:*

72: gb_hc913:*

73: gb_hc914:*

74: gb_hc915:*

75: gb_hc916:*

76: gb_hc917:*

77: gb_hc918:*

78: gb_hc919:*

79: gb_hc920:*

80: gb_hc921:*

81: gb_hc922:*

82: gb_hc923:*

83: gb_hc924:*

84: gb_hc925:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_ro1:*

95: gb_ro2:*

96: gb_in4:*

97: gb_pi10:*

98: em_ba3:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	494	97.6	164371	80	AL390239	AL390239 Homo sapi
2	494	97.6	173831	80	AL359314	AL359314 Homo sapi
3	494	97.6	267581	80	AL358573	AL358573 Homo sapi
4	482	95.3	5435	88	AF103908	AF103908 Homo sapi
5	243.4	48.1	3923	88	AF103907	AF103907 Homo sapi
6	58.2	11.5	7218	10	166494	166494 Sequence 14
7	52.4	10.4	359	9	AX018075	AX018075 Sequence
8	43	8.5	102200	92	HS30M3	AT031775 Human DNA

```

9      42      8.3 215103 72 AC058821 Homo sapi
10     41.2    8.1 9070 13 AF148542 Arabidops
11     41      8.1 175391 69 AC025622 Mus muscu
12     40.8    8.1 132887 90 AL160408 Human DNA
13     40      7.9 199103 76 AC080076 Homo sapi
14     39.8    7.9 166774 85 AC005225 Homo sapi
15     39.4    7.8 127091 61 AC009843 Drosophi1
16     39.4    7.8 156736 86 AC006502 Homo sapi
17     39.4    7.8 164331 87 AC016254 Homo sapi
18     39.4    7.8 166300 66 AC021373 Homo sapi
19     39.4    7.8 171419 86 AC006961 Homo sapi
20     39.4    7.8 191148 70 AC027588 Homo sapi
21     39.4    7.8 222895 76 AC084065 Mus muscu
22     39.2    7.7 160910 83 AP001929 Homo sapi
23     39.2    7.7 184800 75 AC074144 Mus muscu
24     39      7.7 209346 70 AC026373 Homo sapi
25     38.8    7.7 17609 5 AF208112 Trypanoso
26     38.8    7.7 118030 82 AP000667 Homo sapi
27     38.8    7.7 158759 63 AC013484 Trypanoso
28     38.8    7.7 164587 67 AC023001 Homo sapi
29     38.8    7.7 196361 85 AC005386 ciltb_57_1
30     38.8    7.7 203407 60 AC006174 Homo sapi
31     38.6    7.6 88782 67 AC022625 Homo sapi
32     38.4    7.6 159350 68 AC024332 Homo sapi
33     38.4    7.6 200125 71 AC034105 Homo sapi
34     38.2    7.5 150748 68 AC024096 Homo sapi
35     38.2    7.5 164485 69 AC026106 Homo sapi
36     38.2    7.5 171154 83 AP003356 Homo sapi
37     38.2    7.5 192465 83 AP002851 Homo sapi
38     38      7.5 153342 82 AL589949 Homo sapi
39     38      7.5 158373 73 AC068912 Mus muscu
40     38      7.5 236406 76 AC079489 Mus muscu
41     37.8    7.5 2292 8 D14316 D14316 delta-cryst
42     42      7.5 19479 94 MMA21XCOA 222923 M.musculus
43     37.8    7.5 134210 85 AC005052 Homo sapi
44     37.8    7.5 159818 60 AC007337 Homo sapi
45     37.8    7.5 163229 87 AC009039 Homo sapi

```

ALIGNMENTS

```

RESULT 1
AL390239 164371 bp DNA HTG 09-MAR-2001
LOCUS Homo sapiens chromosome 9 clone RP11-58J3, *** SEQUENCING IN
DEFINITION
ACCESSION AL390239.11 GI:13274794
VERSION Homo sapiens
KEYWORDS HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL 1 (bases 1 to 164371)

```

```

COMMENT
-----
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
Center project name: Ba58J3
-----
Summary Statistics
Assembly Program: XGAP4; Version 4.5
Sequencing vector: plasmid: 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 16337 bases at least Q40

```

```

Consensus quality: 163481 bases at least Q30
Consensus quality: 163577 bases at least Q20
Insert size: 164071; sum-of-contigs
Insert size: 166918; 4.28 error; agarose-fp
Quality coverage: 8.78x in Q20 bases; sum-of-contigs Quality
coverage: 9.07x in Q20 bases; agarose-fp
-----

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 20241 20340: gap of 100 bp
* 20341 116100: contig of 95760 bp in length
* 116101 116200: gap of 100 bp
* 116201 118243: contig of 2043 bp in length
* 118244 118343: gap of 100 bp
* 118344 164371: contig of 46028 bp in length.
location/Qualifiers

```

FEATURES

source

```

1..164371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-58J3"
/clone_1lb="RPC1-11.1"
1..20240
/note="assembly-fragment:00838
fragment_chain:1
clone_end:SP6
vector_side:left"
20341..116100
/note="assembly-fragment:01998
fragment_chain:1"
116201..118243
/note="assembly-fragment:00916
fragment_chain:1"
118344..164371
/note="assembly-fragment:00334
fragment_chain:1
clone_end:T7
vector_side:right"

```

```

misc-feature
misc-feature
misc-feature

```

```

misc-feature

```

```

BASE COUNT 50254 a 34905 c 32663 g 46249 t 300 others
ORIGIN

```

```

Query Match 97.6%; Score 494; DB 80; Length 164371;
Best Local Similarity 99.8%; Pred. No. 1,2e-134;
Matches 503; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

```

OY 1 caggaaagacacaaaggaagacacagagtaagtgtcttataaagactcoatttactca 60
Db 33067 CAGGAAAGACACAAAGGAAGACACAGAGTAAGTGTCTTATAAGACACCAATTTCTACTCA
OY 61 gaattttgtgtgcctaaattctctaccgcgtttctatcctctactactactgtcttc 120
Db 33127 GAAATTTTGTGATGGCCTTAAAGTCTTCTACGTTTCTATCTTCTCTACTCTGCTCTC
OY 121 ccggaatccactaccgattctctattctctccgcatgtatgtctgaactgtcacttga 180
Db 33187 CCGGAATCCATACCGATTCTTCTATTTCTTTCGTCGTAATTTCTGACAGGCTCACTTGA
OY 181 ttatctctacagagtgctgaatttctaccgggtcactccgtccctccatattgtc 240
Db 33247 TTTATTCCTACAGAGTCTGAGATTCTTCTACCGGCTCACCTCCCTCCATATTTGTGTC
OY 241 ctccacttccagagatccctgtggagaaatgcgcggccgacatcttggatgatgagc 300
Db 33307 CTCACATTTCACAGATCCCTCTGGAGAAATGCCGCGCCCATCTTGGGTGATTCATGATGAC

```

QY 301 ctcgcccctgtgctgtccgcctgtgtgaggaagagacattagaanaatgatgtgtt 360
|||||
Db 33367 CTCGCCCTGTGCTGTGCTGCCGCTGTGAGGAGAGCATTAAGAAATGATTCATCTGTT 33426
QY 361 ccttaaggaatgaggaagaaacagatcctgtgtgtgatatattttaacggattaca 420
|||||
Db 33427 CCTTAAGGATGGGCGAGAAACAGATCCTGTGTGATATTATTATTTGAACGGATTACA 33486
QY 421 gattgaatgaatgaatgacccaagtgcattaccatgaggaagaaacagagaaat 480
|||||
Db 33487 GATTGAAATGAAGTCA-CAAAGTGAGCATTTACCAATGAGAGGAAACAGACGAGAAAT 33545
QY 481 cttgatgcttcacaagacatgcaac 506
|||||
Db 33546 CTTGATGGCTTCACAAGACATGCAAC 33571
RESULT 2
AL359314 173831 bp DNA HTG 08-APR-2001
LOCUS Homo sapiens chromosome 9 clone RP11-1081L4, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.
ACCESSION AL359314
VERSION AL359314.12 GI:13396560
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 173831)
Direct Submission
Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:13277120.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: bAI081L4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 173545 bases at least Q40
Consensus quality: 173599 bases at least Q30
Consensus quality: 173621 bases at least Q20
Insert size: 173631; sum-of-contrigs
Insert size: 172123; 10.0% error; agarose-fp
Quality coverage: 10.35x in Q20 bases; sum-of-contrigs Quality
coverage: 10.50x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 50595: contig of 50595 bp in length
* 50596 50695: gap of 100 bp
* 50696 157636: contig of 106941 bp in length
* 157637 157736: gap of 100 bp
* 157737 173831: contig of 16095 bp in length.
Location/Qualifiers
1. 173831
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"

/clone="RP11-1081L4"
/clone.lib="RPC1-11.1"
1. 50595
/note="assembly-fragment:04137
fragment_chain:1"
50696. 157636
/note="assembly-fragment:04236
fragment_chain:1"
157737. 173831
/note="assembly-fragment:01566"
BASE COUNT 52107 a 36469 c 35018 g 50037 t 200 others
ORIGIN
Query Match 97.6%; Score 494; DB 80; Length 173831;
Best Local Similarity 99.8%; Pred. No. 1.2e-134;
Matches 505; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1 caggaagacaaagaagacacagagtaagtcttataaagcactcaattctactca 60
|||||
Db 155988 CAGGAAGCACAAAGAGACAGACAGAGTAAGTCTTTATRAAGCACTCAATTTCTACTCA 156047
QY 61 gaaattttgatggccttaagtccctactactcgtttctactactactactactcctc 120
|||||
Db 156048 GAATTTTGTGATGGCCTTAAGTTCCTACTCTGTTCTATTCCTTCTACTACTGTCCTC 156107
QY 121 ccggaatccactacagatttctatttctgtcgttatgtctgactgctcacttgga 180
|||||
Db 156108 CCGGAATCCACTCCAGATTTTCATTTCTTGCTCCTCATTTGTCAGTGGCTCACTTGA 156167
QY 181 ttatctctacagagctgctgatttctaccgggctcacctccgtccctccatatttgc 240
|||||
Db 156168 TTTATCTTCACGAGATCTGTGATTTTCTACCCGGGCTCACCTCCCTCCCAATTTGTTC 156227
QY 241 ctcacattcacagatccctgaggaagaaatgcccgcgcacatcttggtcatcgatgaac 300
|||||
Db 156228 CTCACATTTCACAGATCCCTGGAGAAATCCCGCGCCGACATCTTGCGATGATGAGAC 156287
QY 301 ctcgcccctgtgctgtccgcctgtgtgaggaagagacattagaanaatgatgtgtt 360
|||||
Db 156288 CTCGCCCTGTGCTGTGCTGCCGCTGTGAGGAGAGCATTAGAAATGATTCATCTGTT 156347
QY 361 ccttaaggaatgaggaagaaacagatcctgtgtgtgatatattttaacggattaca 420
|||||
Db 156348 CCTTAAGGATGGGCGAGAAACAGATCCTGTGTGATATTATTATTTGAACGGATTACA 156407
QY 421 gattgaatgaatgaatgacccaagtgcattaccatgaggaagaaacagagaaat 480
|||||
Db 156408 GATTGAAATGAAGTCA-CAAAGTGAGCATTTACCAATGAGAGGAAACAGACGAGAAAT 156466
QY 481 cttgatgcttcacaagacatgcaac 506
|||||
Db 156467 CTTGATGGCTTCACAAGACATGCAAC 156492
RESULT 3
AL358573/c 267581 bp DNA HTG 15-APR-2001
LOCUS Homo sapiens chromosome 9 clone RP11-133022, *** SEQUENCING IN
DEFINITION PROGRESS ***, 37 unordered pieces.
ACCESSION AL358573
VERSION AL358573.17 GI:13660951
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 267581)
Direct Submission
Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

COMMENT

On Apr 17, 2001 this sequence version replaced g1:13398774.
----- Genome Center

Center: Sanger Centre

Center code: SC

*Web site: <http://www.sanger.ac.uk>

Contact: humquerry@sanger.ac.uk

----- Project Information

Center project name: BA133022

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 256699 bases at least Q40

Consensus quality: 259744 bases at least Q30

Consensus quality: 261407 bases at least Q20

Insert size: 263981; sum-of-coverage

Insert size: 135491; 19.3% error; agarose-fp

Quality coverage: 5.5ix in Q20 bases; sum-of-coverage

coverage: 12.23x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* 1 25718: contig of 25718 bp in length
* 25719 25818: gap of 100 bp
* 25819 29567: contig of 3749 bp in length
* 29568 29667: gap of 100 bp
* 29668 40520: contig of 10853 bp in length
* 40521 40620: gap of 100 bp
* 40621 50002: contig of 9382 bp in length
* 50003 50102: gap of 100 bp
* 50103 54942: contig of 4840 bp in length
* 54943 55042: gap of 100 bp
* 55043 59067: contig of 4025 bp in length
* 59068 59167: gap of 100 bp
* 59168 72327: contig of 13160 bp in length
* 72328 72427: gap of 100 bp
* 72428 79396: contig of 6969 bp in length
* 79397 79496: gap of 100 bp
* 79497 82738: contig of 3242 bp in length
* 82739 82838: gap of 100 bp
* 82839 87543: contig of 4705 bp in length
* 87544 87643: gap of 100 bp
* 87644 89847: contig of 2204 bp in length
* 89848 89947: gap of 100 bp
* 89948 92376: contig of 2429 bp in length
* 92377 92476: gap of 100 bp
* 92477 99799: contig of 7223 bp in length
* 99800 99899: gap of 100 bp
* 99900 102117: contig of 2218 bp in length
* 102118 102217: gap of 100 bp
* 102218 109330: contig of 7113 bp in length
* 109331 109430: gap of 100 bp
* 109431 112187: contig of 2757 bp in length
* 112188 112287: gap of 100 bp
* 112288 116407: contig of 4120 bp in length
* 116408 116507: gap of 100 bp
* 116508 119386: contig of 2879 bp in length
* 119387 119486: gap of 100 bp
* 119487 121869: contig of 2283 bp in length
* 121870 121969: gap of 100 bp
* 121970 125546: contig of 3577 bp in length
* 125547 125646: gap of 100 bp
* 125647 128990: contig of 3344 bp in length
* 128991 129090: gap of 100 bp
* 129091 134055: contig of 4965 bp in length
* 134056 134155: gap of 100 bp
* 134156 138314: contig of 4159 bp in length

```

FEATURES

source

```

* 138315 138414: gap of 100 bp
* 138415 140612: contig of 2196 bp in length
* 140613 140712: gap of 100 bp
* 140713 143296: contig of 2584 bp in length
* 143297 143396: gap of 100 bp
* 143397 148113: contig of 4717 bp in length
* 148114 148213: gap of 100 bp
* 148214 151009: contig of 2796 bp in length
* 151010 151109: gap of 100 bp
* 151110 156616: contig of 5507 bp in length
* 156617 156716: gap of 100 bp
* 156717 158763: contig of 2047 bp in length
* 158764 158863: gap of 100 bp
* 158864 162129: contig of 3266 bp in length
* 162130 162229: gap of 100 bp
* 162230 164443: contig of 2214 bp in length
* 164444 164543: gap of 100 bp
* 164544 169800: contig of 5257 bp in length
* 169801 169900: gap of 100 bp
* 169901 2251545: contig of 81645 bp in length
* 2251546 251645: gap of 100 bp
* 251646 259029: contig of 7384 bp in length
* 259030 259129: gap of 100 bp
* 259130 262871: contig of 3742 bp in length
* 262872 262971: gap of 100 bp
* 262972 265209: contig of 2238 bp in length
* 265210 265309: gap of 100 bp
* 265310 267581: contig of 2272 bp in length.

```

Location/Qualifiers

```

1. .267581
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="9"
  /clone="RP11-133022"
  /clone_1bp="RP11-11.1"
  1. .25718
    /note="assembly_fragment:00645"
    fragment_chain:1
      clone_end:T7
      vector_side:left"
      25819. .29567
        /note="assembly_fragment:01660"
        fragment_chain:1"
      29668. .40520
        /note="assembly_fragment:03080"
        fragment_chain:1"
      40621. .50002
        /note="assembly_fragment:00771"
        fragment_chain:1"
      50103. .54942
        /note="assembly_fragment:02629"
        fragment_chain:1"
      55043. .59067
        /note="assembly_fragment:04565"
        fragment_chain:1"
      59168. .72327
        /note="assembly_fragment:00223"
        fragment_chain:1"
      72428. .79396
        /note="assembly_fragment:03318"
        fragment_chain:1"
      79497. .82738
        /note="assembly_fragment:00224"
        fragment_chain:2"
      82839. .87543
        /note="assembly_fragment:03050"
        fragment_chain:2"
      87644. .89847
        /note="assembly_fragment:01948"
        fragment_chain:3"
      89948. .92376
        /note="assembly_fragment:02191"
        fragment_chain:3"

```

misc_feature	92477. .99799
misc_feature	/note="assembly_fragment:00010"
misc_feature	99900. .102117
misc_feature	/note="assembly_fragment:00672"
misc_feature	102218. .109330
misc_feature	/note="assembly_fragment:01326"
misc_feature	109431. .112187
misc_feature	/note="assembly_fragment:01498"
misc_feature	112288. .116407
misc_feature	/note="assembly_fragment:01607"
misc_feature	116508. .119386
misc_feature	/note="assembly_fragment:01699"
misc_feature	119487. .121869
misc_feature	/note="assembly_fragment:02068"
misc_feature	121970. .125546
misc_feature	/note="assembly_fragment:02469"
misc_feature	125647. .128990
misc_feature	/note="assembly_fragment:02502"
misc_feature	129091. .134055
misc_feature	/note="assembly_fragment:02669"
misc_feature	134156. .138314
misc_feature	/note="assembly_fragment:02757"
misc_feature	138415. .140612
misc_feature	/note="assembly_fragment:02776"
misc_feature	140713. .143296
misc_feature	/note="assembly_fragment:03127"
misc_feature	143397. .148113
misc_feature	/note="assembly_fragment:03634"
misc_feature	148214. .151009
misc_feature	/note="assembly_fragment:03898"
misc_feature	151110. .156616
misc_feature	/note="assembly_fragment:04220"
misc_feature	156717. .156763
misc_feature	/note="assembly_fragment:04276"
misc_feature	158864. .162129
misc_feature	/note="assembly_fragment:04418"
misc_feature	162250. .164443
misc_feature	/note="assembly_fragment:04428"

Query Match	97.6%	Score 494	DB 80	Length 267581
Best Local Similarity	99.8%	Pred. No. 1.3e-134		
Matches 505	Conservative 0	Mismatches 0	Indels 1	Gaps 1
QY 1	caggaagcacaaaaggaagcacagagatgaagtgtccttaataagcactcaattctactca	60		
Dp 166149	CAGGAGACGCAAAAAGGAAGCACAGAGGTAAGTGCTTTTAAAGCATCAATTTCCTAC	166090		
QY 61	gaatttttgatgagccttaagtctccctactcgtttctatccctctactcaatcagtcctc	120		
Dp 166089	GAAATTTTGTATGGCCCTTAAGTCCCTACACGCTTTCATGCTTCCTACATCAGTCCGC	166030		
QY 121	cgggaatccactacacgaattttctatctttcttcctcgtatgtttctacagcgccacttga	180		
Dp 166029	CCGGAAATCCACTACCGAATTTCTATTTCTTCTTCCTCGATGCTATGCTACGGCTCACTTGA	165970		
QY 181	tttaacctcacagagatcttgaatttctaacccgggctcacctccgcgtccctccatattgct	240		
Dp 165969	TTTATCCGACAGAGATGTCGATTTTCTATCCGGGGCTACCTCCGCTCCATATTGTGC	165910		
QY 241	ctccacttccacagatcccttggagagaatgcccggcgccacatcttggatccatgatgagc	300		
Dp 165909	CTCCACTTTTCACAATATCCCTCTGGAGAAATGCCCGGCCGACCATCTTGGGTCATCGATGAGC	165850		
QY 301	ctggcccccgtgcccgtgctccgctgttggagggaagacattgaaaataatgatgagtgt	360		
Dp 165849	CTCGCCCTGTGCTCCCTGTGCCCTGTGTGAGGAAGGACATTTGAAAATGATGATGATGT	165790		
QY 361	ccttaaaaggaatgaggcaggaanaacagatcctgttctgataattatatttgaacggagattaca	420		
Dp 165789	CCTTAAAGGATGGGACAGSAAAACAGATCCTGTGTGGATATTATTATTGAACGGATTTACA	165730		
QY 421	gatttgaatgaagtccaccaagtgaagcattaccaatgagaggaacacagacgagaaat	480		

	QY	481	cttgatggcttcacagacatgcgaac	506	
	Db	165670	CTTGATGGCTTCACAGACATGCGAAC	165645	
	RESULT	4		PRI	14-AUG-2000
	LOCUS	AFI03908	AFI03908	5435 bp	DNA
	DEFINITION	Homo sapiens non-coding RNA	DD3 gene,	exons 2, 3,	and 4.
	ACCESSION	AFI03908			
	VERSION	AFI03908.1	GI:6165974		
	KEYWORDS	.			
	SOURCE	human.			
	ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
	REFERENCE	1 (bases 1 to 5435) Bussemakers,M.J., van Bokhoven,A., Verhaegh,G.W., Smit,F.P., Karthaus,H.F., Schalken,J.A., Debruyne,F.M., Ru,N. and Isaacs,W.B. DD3 : a new prostate-specific gene, highly overexpressed in prostate cancer			
	AUTHORS	Cancer Res. 59 (23), 5975-5979 (1999)			
	JOURNAL	20072260			
	MEDLINE	10606244			
	PUBMED	2 (bases 1 to 5435)			
	REFERENCE	Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P., Karthaus,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru,N. and Isaacs,W.B.			
	AUTHORS	Direct Submission Submitted (28-OCT-1998) Urology Research Laboratory, University Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands			
TITLE	JOURNAL	Location/Qualifiers			
FEATURES	source	1. .5435 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="9" /map="q21-q22"			
		1. .78 /rpt_family="Alu"			
	repeat_region	/rpt_type-dispersed 533..697 /note="alternative exon present in 5% of cDNA clones"			
	excn	/number=2 1035..1294 /rpt_family="Alu"			
	repeat_region	/rpt_type-dispersed join(1571..1753,1981..5435) /note="transcript II"			
	mRNA	/product="non-coding RNA DD3" join(1571..1753,1981..2517) /note="transcript I"			
	mRNA	/product="non-coding RNA DD3" join(1571..1753,1981..3579) /note="transcript (major) II"			
	excn	/product="non-coding RNA DD3" 1571..1753 /number=3 1981..5435			
	excn	/number=4 2495..2499 /note="transcript I"			
	polyA_signal	2517 /note="transcript I"			
	polyA_site	3553..3558 /note="transcript (major) II"			
	polyA_signal	3579 /note="transcript (major) II"			
	polyA_site	5268..5423 /note="LINE"			
	repeat_region	/rpt_family="L1" /rpt_type-dispersed			

[illegible]

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence is the entire insert of clone 30M3. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>

30M3 is from the library RPCI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcrPAC2>.

Location/Qualifiers
1..102200

UNES	
source	Location/Qualifiers
	1..102200 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="6" map="p22.1-22.3" /clone="Rpl-30M3" /clone_1lib="RPlC1-1"
repeat_region	757..825 /note="MER47 repeat: matches 9..76 of consensus"
repeat_region	829..932 /note="MER47 repeat: matches 2226..2319 of consensus"
repeat_region	905..976 /note="MER47 repeat: matches 2249..2323 of consensus"
repeat_region	1320..1641 /note="AluSc repeat: matches 1..299 of consensus"
repeat_region	1642..1927 /note="AluDb repeat: matches 1..289 of consensus"
repeat_region	2392..2517 /note="FLAM_C repeat: matches 1..126 of consensus"
repeat_region	2760..2789 /note="I5 copies 2 mer tg 97% conserved"
repeat_region	3013..3313 /note="AluDo repeat: matches 1..301 of consensus"
repeat_region	3467..3763 /note="AluSp repeat: matches 1..299 of consensus"
misc_feature	complement(3650..4203) /note="match: GSS B63434"
repeat_region	4458..4591 /note="MIR repeat: matches 53..208 of consensus"
repeat_region	4799..4952 /note="MER3 repeat: matches 3..154 of consensus"
repeat_region	join<5119..5253,7768..9169)
RNA	/gene="dJ30M3.2"
	/note="match: ESTs AA259531 AA289775 N50863 AA806198 AA931087 AA830274 AI199746 AA912273 AT081164 A1288433 AA604069 AA772465 AA876636 AA521057 A1147796 AA639167 AA665259 AA289775 AA259531 AA731052 AA290469 D20450 AA517053 AA669103 AA974142 AA903560 AT083595 A1287615 A1041875 A1234339 AA5220 R23494 AA798467 AA798123 AA551228 AA470882 A1210905 AA663774 AA840219 AA041897 D77719 AT019173 AT006885 A1194277 D56188 AA926179 AA851214 T519377 R01213: supported by GENSCAN"
gene	/evidence=not_experimental /product="dJ30M3.2 (novel protein)" 5119..9169 /gene="dJ30M3.2"
CDS	join<5119..5253,7768..7893) /gene="dJ30M3.2" /note="supported by GENSCAN" /codon_start=1 /evidence=not_experimental /product="dJ30M3.2 (novel protein)"

```

/protein_id="CAA21140.1"
/db_xref="GI:4210352"
/translation="AKFEFHGDYKOPHLVSRDKTGIVVNNQSVFLEIDNQH
LQTPKKAITFKICSLTLPQQLTHMAGTIEDLRLPMP"
repeat_region
  /note="AluSg repeat: matches 1. .291 of consensus"
  5393. .5674
repeat_region
  /note="MER5B repeat: matches 1. .49 of consensus"
  5715. .5763
repeat_region
  /note="MER5B repeat: matches 1. .189 of consensus"
  5883. 6000
repeat_region
  /note="AluY repeat: matches 1. .295 of consensus"
  6010. 6308
repeat_region
  /note="AluY repeat: matches 1. .296 of consensus"
  6384. .6667
repeat_region
  /note="AluJo repeat: matches 5. .207 of consensus"
  6740. .6878
repeat_region
  /note="MIR repeat: matches 69. .207 of consensus"
  7106. .7400
repeat_region
  /note="AluJb repeat: matches 1. .297 of consensus"
  8676. .8721
repeat_region
  /note="23 copies 2 mer tt 78% conserved"
  9092. .9129
repeat_region
  /note="19 copies 2 mer aa 82% conserved"
  9124. .9129
polyA_signal
  /gene="dJ30M3.2"
  9169
polyA_site
  /gene="dJ30M3.2"
  9439. .9621
repeat_region
  /note="MIR repeat: matches 8. .256 of consensus"
  9691. .9798
repeat_region
  /note="L1MC4 repeat: matches 7851. .7977 of consensus"
  10020. .10149
repeat_region
  /note="L1MD3 repeat: matches 7361. .7499 of consensus"
  10507. .10716
repeat_region
  /note="L1MC4 repeat: matches 6633. .6858 of consensus"
  10855. .11115
repeat_region
  /note="AluJo repeat: matches 1. .283 of consensus"
  11469. .11619
repeat_region
  /note="MIR1A1 repeat: matches 1. .146 of consensus"
  11620. .11806
repeat_region
  /note="AluSx repeat: matches 1. .287 of consensus"
  11907. .12055
repeat_region
  /note="MIR1A1 repeat: matches 146. .310 of consensus"
  complement(join(12316. .12571,15963. .>16147))
  /gene="dJ30M3.1"
  /note="mach: ESTs AA5222110 AA207689 AA819653 AA928745
  AI017897 AA993303 AI216348 AA832106 AA121362 AA552456
  AA629886 AA614605 T55234 AA121362 F17737 AA445103 AA073807
  AA121372 AI052396 AA939005 AI267851 T53974; supported by
  GENSCAN and FGENES"
  /evidence="not_experimental"
  /product="dJ30M3.1 (novel protein similar to (predicted)
  plant, worm, yeast and archaea bacterial proteins)"
  complement(12316)
polyA_site
  /gene="dJ30M3.1"
  complement(12316. .16147)
  /gene="dJ30M3.1"
  complement(12331. .12336)
  /gene="dJ30M3.1"
  complement(join(12415. .12571,15963. .>16147))
  /gene="dJ30M3.1"
  /note="match: proteins P87304 P93828 CE02501 CE00158
  P93914; supported by GENSCAN and FGENES"
  /codon_start=1
  /evidence="not_experimental"
  /product="dJ30M3.1 (novel protein similar to (predicted)
  plant, worm, yeast and archaea bacterial proteins)"
  /protein_id="CAA21139.1"
  /db_xref="GI:4210351"
  /db_xref="GI:4210351"
  /db_xref="SPTREMBL:095549"
  /translation="ITLVASAPKGVICEMKVEEHTNATIGLHGLATLVDNISTMA
  LITCRGAPGVSYDMNITVMSPAKLGEIDIVITAHVLNQGKTLAFTSDVLTAKAKGKLI
  AOGHRTKHLGN"

```

```

repeat_region
  13295. .13604
  /note="AluSx repeat: matches 1. .310 of consensus"
  14521. .14821
  /note="AluYb8 repeat: matches 1. .299 of consensus"
  15123. .15387
  /note="AluSx repeat: matches 30. .298 of consensus"
  15514. .15624
  /note="AluJo repeat: matches 2. .136 of consensus"
  16957. .17250
  /note="AluSg repeat: matches 1. .294 of consensus"
  17526. .17589
  /note="MIR repeat: matches 194. .258 of consensus"
  17534. .17637
  /note="L2 repeat: matches 2588. .2692 of consensus"
  17943. .18245
  /note="AluYas5 repeat: matches 1. .311 of consensus"
  18551. .18684
  /note="L2 repeat: matches 2357. .2495 of consensus"
  18704. .18998
  /note="AluSx repeat: matches 6. .300 of consensus"
  19037. .19404
  /note="match: GSS B40022"
  19504. .19639
  /note="MIR repeat: matches 88. .238 of consensus"
  20030. .20303
  /note="AluSx repeat: matches 24. .297 of consensus"
  complement(20331. .20463)
  /note="match: GSS B74972"
  21207. .21337
  /note="MIR repeat: matches 90. .244 of consensus"
  21347. .21600
  /note="AluJo repeat: matches 36. .287 of consensus"
  21692. .21891
  /note="L2 repeat: matches 1320. .1536 of consensus"

Query Match
Best Local Similarity 8.5%; Score 43; DB 92; Length 102200;
Matches 94; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 66 ttltgatgccttaagttctctactcgttctatcctcctcactcactgctccgga 125
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71711 TTTAAATCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTC 71770

Qy 126 atccactcagatcttctattctctgctcgtatgctgactgctcactgattat 185
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71771 CCTTCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC 71830

Qy 186 cctcagcagatctgatttctcaccgagctcactcgtccctccatattgctccc 244
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71831 TTTCTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTCC 71889

RESULT 9
AC058821/c DNA HTG 24-FEB-2001
LOCUS AC058821 215103 bp
DEFINITION Homo sapiens chromosome 4 clone RP11-778B12, WORKING DRAFT
ACCESSION AC058821
VERSION AC058821.2 GI:7767912
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 215103)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 215103)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

```


COMMENT

MO 63108, USA
On May 10, 2000 this sequence version replaced gi:7596990.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H_NH0778812
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 189194 bases at least Q40
Consensus quality: 197110 bases at least Q30
Consensus quality: 201125 bases at least Q20
Insert size: 207000; agarose-fp
Quality coverage: 3.08 in Q20 bases; agarose-fp
Quality coverage: 3.01 in Q20 bases; sum-of-ctgigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 46 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1190: contig of 1190 bp in length
* 1191 1280: gap of unknown length
* 1291 2491: contig of 1201 bp in length
* 2492 2591: gap of unknown length
* 2592 3970: contig of 1379 bp in length
* 3971 4070: gap of unknown length
* 4071 6259: contig of 2189 bp in length
* 6260 6359: gap of unknown length
* 6360 8648: contig of 2289 bp in length
* 8649 8749 10583: contig of 1835 bp in length
* 8749 10584 10683: gap of unknown length
* 10684 13046: contig of 2363 bp in length
* 13047 13146: gap of unknown length
* 13147 15724: contig of 2578 bp in length
* 15725 15824: gap of unknown length
* 15825 18468: contig of 2644 bp in length
* 18469 18568: gap of unknown length
* 18569 21077: contig of 2509 bp in length
* 21078 21177: gap of unknown length
* 21178 24304: contig of 3127 bp in length
* 24305 24404: gap of unknown length
* 24405 27866: contig of 3462 bp in length
* 27867 27966: gap of unknown length
* 27967 31186: contig of 3220 bp in length
* 31187 31286: gap of unknown length
* 31287 34406: contig of 3120 bp in length
* 34407 34506: gap of unknown length
* 34507 36959: contig of 2453 bp in length
* 36960 37060: gap of unknown length
* 37061 40266: contig of 3207 bp in length
* 40267 40366: gap of unknown length
* 40367 43164: contig of 2798 bp in length
* 43165 43264: gap of unknown length
* 43265 46349: contig of 3085 bp in length
* 46350 46449: gap of unknown length
* 46450 49752: contig of 3303 bp in length
* 49753 49852: gap of unknown length
* 49853 52949: contig of 3097 bp in length
* 52950 55737: contig of 2688 bp in length
* 55738 55837: gap of unknown length

FEATURES

source

55838 58954: contig of 3117 bp in length
* 58955 59054: gap of unknown length
* 59055 62369: contig of 3315 bp in length
* 62370 62469: gap of unknown length
* 62470 65962: contig of 3493 bp in length
* 65963 66062: gap of unknown length
* 66063 70029: contig of 3967 bp in length
* 70030 70129: gap of unknown length
* 70130 73918: contig of 3789 bp in length
* 73919 74018: gap of unknown length
* 74019 78049: contig of 4031 bp in length
* 78050 78149: gap of unknown length
* 78150 84568: contig of 6419 bp in length
* 84569 84668: gap of unknown length
* 84669 91351: contig of 6683 bp in length
* 91352 91451: gap of unknown length
* 91452 96851: contig of 5400 bp in length
* 96852 96951: gap of unknown length
* 96952 102096: contig of 5145 bp in length
* 102097 102196: gap of unknown length
* 102197 107711: contig of 5515 bp in length
* 107712 107811: gap of unknown length
* 107812 111824: contig of 4013 bp in length
* 111825 111924: gap of unknown length
* 111925 116715: contig of 4791 bp in length
* 116716 116815: gap of unknown length
* 116816 121462: contig of 4647 bp in length
* 121463 121562: gap of unknown length
* 121563 127640: contig of 6078 bp in length
* 127641 127740: gap of unknown length
* 127741 133093: contig of 5353 bp in length
* 133094 133193: gap of unknown length
* 133194 139965: contig of 6772 bp in length
* 139966 140065: gap of unknown length
* 140066 146801: contig of 6736 bp in length
* 146802 146901: gap of unknown length
* 146902 154099: contig of 7198 bp in length
* 154100 154199: gap of unknown length
* 154200 163123: contig of 8924 bp in length
* 163124 163223: gap of unknown length
* 163224 171228: contig of 8005 bp in length
* 171229 171328: gap of unknown length
* 171329 179017: contig of 7689 bp in length
* 179018 179117: gap of unknown length
* 179118 188418: contig of 9301 bp in length
* 188419 188518: gap of unknown length
* 188519 199297: contig of 10779 bp in length
* 199298 199397: gap of unknown length
* 199398 215103: contig of 15706 bp in length.
Location/Qualifiers
1. 215103
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-778B12"
1. 1190
/note="assembly_name:Contig13"
1291. 2491
/note="assembly_name:Contig18"
2592. 3970
/note="assembly_name:Contig19"
4071. 6259
/note="assembly_name:Contig20"
6360. 8648
/note="assembly_name:Contig21"
8749. 10583
/note="assembly_name:Contig22"
10684. 13046
/note="assembly_name:Contig23"
13147. 15724
/note="assembly_name:Contig25"
15825. 18468
/note="assembly_name:Contig26"

Matches	44; Conservative	160; Mismatches	156; Indels	0; Gaps	0; Duplicates
QY 54	ctaccagaattttgttggtcctaagtctctctactcgtttctatctctctactc	113			
Db 8951	MDYHYHHKNNBCATWARYBYBVMYHYUSKMBHYKKKKYUHSNDKNCYTCC	8892			
QY 114	ttgtccccgaataccataccagattttctatttctctcgtatcttctgaactg	173			
Db 8891	CCBSRSSNCTASCMWRSRSYASSMANBTRCARCRNGTBNSTHMYHYSSTRSCY	8832			
QY 174	actggatttctcccaaggagtcgatttcttaccggcggtctaacctccgctcc	233			
Db 8831	YHTSNWTKGMBBYTSRNGSBCTDCTKBNBNTWISCAACTYRMBAIYMYADYK	8772			
QY 234	attgtctccacatttccacagatccctctggagaatgcccgccgacattctgg	293			
Db 8771	HMCRRDYTMVDBCTCBVSTBRTYHNAYSBMAIYTHGKMHNTBSSIBYAMB	8712			
QY 294	gatagcctccgcctgtgctctgttccgcgcttctgtagggagaagacattga	353			
Db 8711	CGYBSSSTYMWMCYTHCRBAYVSTFSBYADPTTYNRRSGAAGRHNGKCSY	8652			
QY 354	atgtgtccttaaggaagagcgacgaagaacagatccctgttctgatatattt	413			
Db 8651	ASSHSSSAYYAASDCSKBYAYCARCABAISNCHSSNSBSYNSAAYADTKA	8592			
RESULT 11					
AC025622	175391 bp	DNA	HTG	26-AUG-2000	
LOCUS	Mus musculus clone RP23-453M11	WORKING DRAFT SEQUENCE	29		
DEFINITION	unordered pieces.				
ACCESSION	AC025622				
VERSION	GI:9929751				
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	1 (bases 1 to 175391)				
ATTORNS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 175391)				
REFERENCE	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campolongo, A., Castle, A., Chovel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArle, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hegos, B., Hefford, A., Horton, L., Kiehl, J., Klei, J., Klei, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaHocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., MCPhetters, R., Meldrum, J., Menus, L., Mihova, T., Miraneda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tefaye, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT	On Aug 26, 2000 this sequence version replaced gi:7230199. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html				
	Genome Center				

```

Center Whitehead Institute/MIT Center for Genome Research
Center code: MIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L5129
Center clone name: 453_M.11
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 160619 bases at least Q40
Consensus quality: 168125 bases at least Q30
Consensus quality: 170821 bases at least Q20
Insert size: 182000; agarose-ftp
Insert size: 172591; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-ftp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 29 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
2946: contig of 2946 bp in length
2947 3046: gap of 100 bp
3047 4282: contig of 1235 bp in length
4282 4381: gap of 100 bp
4382 5991: contig of 1610 bp in length
5992 6091: gap of 100 bp
6092 21312: contig of 15221 bp in length
21313 21412: gap of 100 bp
21413 23033: contig of 1621 bp in length
23034 23133: gap of 100 bp
23134 26440: contig of 3307 bp in length
26441 26540: gap of 100 bp
26541 31551: contig of 5011 bp in length
31552 31651: gap of 100 bp
31652 36488: contig of 4837 bp in length
36489 36588: gap of 100 bp
36589 40281: contig of 3693 bp in length
40282 40381: gap of 100 bp
40382 43320: contig of 2939 bp in length
43321 43420: gap of 100 bp
43421 49736: contig of 6316 bp in length
49737 49836: gap of 100 bp
49837 55164: contig of 5328 bp in length
55165 55264: gap of 100 bp
55265 60373: contig of 5109 bp in length
60374 60473: gap of 100 bp
60474 66509: contig of 6036 bp in length
66510 66609: gap of 100 bp
66610 71217: contig of 4608 bp in length
71218 71317: gap of 100 bp
71318 77911: contig of 6594 bp in length
77912 78011: gap of 100 bp
78012 85330: contig of 7319 bp in length
85331 85430: gap of 100 bp
85431 91064: contig of 5634 bp in length
91065 91164: gap of 100 bp
91165 97554: contig of 6390 bp in length
97555 97654: gap of 100 bp
97655 103507: contig of 5853 bp in length
103508 103607: gap of 100 bp
103608 111290: contig of 7683 bp in length
111291 111390: gap of 100 bp
111391 117312: contig of 5922 bp in length
117313 117412: gap of 100 bp
117413 127398: contig of 9986 bp in length
127399 127498: gap of 100 bp

```

```

* 127499 136874: contig of 9376 bp in length
* 136875 136974: gap of 100 bp
* 136975 144638: contig of 7664 bp in length
* 144639 144738: gap of 100 bp
* 144739 151643: contig of 6905 bp in length
* 151644 151743: gap of 100 bp
* 151744 161007: contig of 9264 bp in length
* 161008 175136: contig of 14029 bp in length
* 175137 175236: gap of 100 bp
* 175237 175391: contig of 155 bp in length.
FEATURES
    source
        1..175391
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="RP23-453M11"
            /clone_1lb="RPC1-23 Female Mouse BAC"
            1..2946
                /note="assembly-fragment"
                clone_end:SP6
                vector_side:left"
            /note="assembly-fragment"
            4382..5991
                /note="assembly-fragment"
                /note="assembly-fragment"
            6092..21312
                /note="assembly-fragment"
            21413..23033
                /note="assembly-fragment"
            23134..26440
                /note="assembly-fragment"
            26541..31551
                /note="assembly-fragment"
            31652..36488
                /note="assembly-fragment"
            36589..40281
                /note="assembly-fragment"
            40382..43320
                /note="assembly-fragment"
            43421..49736
                /note="assembly-fragment"
            49837..55164
                /note="assembly-fragment"
            55265..60373
                /note="assembly-fragment"
            60474..66509
                /note="assembly-fragment"
            66610..71217
                /note="assembly-fragment"
            71318..77911
                /note="assembly-fragment"
            78012..85330
                /note="assembly-fragment"
            85431..91064
                /note="assembly-fragment"
            91165..97554
                /note="assembly-fragment"
            97655..103507
                /note="assembly-fragment"
            103608..111290
                /note="assembly-fragment"
            111391..117312
                /note="assembly-fragment"
            117413..127398
                /note="assembly-fragment"
            127499..136874
                /note="assembly-fragment"
            136975..144638
                /note="assembly-fragment"
            144739..151643
                /note="assembly-fragment"
            151744..161007
                /note="assembly-fragment"

```

```

misc-feature 161108..175136
              /note="assembly-fragment"
misc-feature 175237..175391
              /note="assembly-fragment"
              clone_end:17
              vector_side:right"
BASE COUNT 48326 a 37575 c 37702 g 48984 t 2804 others
ORIGIN

```

```

Query Match      8.1%; Score 41; DB 69; Length 175391;
Best Local Similarity 47.8%; Pred. No. 0.62;
Matches 119; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

```

```

Qy 32 tgcctataagcaactcaattcttactcagaatttgatggccttaagtccttact 91
    |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db 64836 TTCTTTCTCTTCTCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTTC 64895
    |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Qy 92 cgltctatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctg 151
    |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db 64896 TTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTC 64955
    |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Qy 152 cctcgatgtctgactgctcacttgatattatcctcctcctcctcctcctcctcctc 211
    |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db 64956 TTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTC 65015
    |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Qy 212 gggctcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctg 271
    |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db 65016 TTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTC 65075
    |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Qy 272 ccggccgcgc 280
    |||||  |||  |||||  |||  |||||  |||  |||||  |||  |||||
Db 65076 CCTCCGAC 65084

```

```

RESULT 12
AL160408 132887 bp DNA PRI 25-MAR-2001
LOCUS AL160408/C
DEFINITION Human DNA sequence from clone RP4-781K5 on chromosome 1q42.1-43,
complete sequence.
ACCESSION AL160408
VERSION AL160408.24 GI:13446408
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 132887)
AUTHORS Bates,K.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humgeny@sanger.ac.uk
COMMENT On Mar 25, 2001 this sequence version replaced gi:13274316.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; SW.,
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human

```

chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Ch1>
RP4-781K5 is from the library RPI-4 constructed by the group of
Pleier de Jong. For further details see

<http://www.chori.org/Dacpac/home.htm>
VECTOR: pCYPAC2

This sequence is the entire insert of clone RP4-781K5 The true left
end of clone RPI1-400F1 is at 75536 in this sequence. The true left
end of clone RPI1-380I8 is at 104240 in this sequence. The true
right end of clone RP5-855F14 is at 15526 in this sequence.

FEATURES

```
source
1..132887
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.1-43"
/clone="RP4-781K5"
/clone_lib="RPI-4"
3739..6486
/note="Cpg island"
/evidence="not_experimental"
repeat_region
4673..4765
/note="31 copies 3 mer ggc 65% conserved"
4752..4817
/note="33 copies 2 mer cc 68% conserved"
4816..5031
/note="72 copies 3 mer ggc 59% conserved"
4951..5248
/note="149 copies 2 mer gg 58% conserved"
4961..5235
/note="5 copies 55 mer 65% conserved"
5125..5424
/note="Single clone region. Sequence from reads from a
short insert library derived from a single m13 clone.
Restriction digest data confirm the assembly."
5333..5374
/note="14 copies 3 mer ccg 81% conserved"
10072..10367
/note="Alusg repeat: matches 1..296 of consensus"
11031..11339
/note="Alusx repeat: matches 1..308 of consensus"
12204..12235
/note="8 copies 4 mer tacc 90% conserved"
12993..13160
/note="84 copies 2 mer at 75% conserved"
13947..13982
/note="18 copies 2 mer ac 83% conserved"
15995..16181
/note="Aluud repeat: matches 1..167 of consensus"
16181..16389
/note="Aluub repeat: matches 112..310 of consensus"
17813..18003
/note="Tigerr2a repeat: matches 1..191 of consensus"
18004..18297
/note="Alusg repeat: matches 1..306 of consensus"
18298..18539
/note="Tigerr2a repeat: matches 190..434 of consensus"
19058..19133
/note="19 copies 4 mer acac 72% conserved"
19686..19940
/note="MIR repeat: matches 2..261 of consensus"
20066..20186
/note="L2 repeat: matches 1304..1432 of consensus"
20319..20364
/note="L2 repeat: matches 2699..2744 of consensus"
20532..20840
/note="Aluub repeat: matches 1..307 of consensus"
21075..21236
/note="FRAM repeat: matches 4..165 of consensus"
21242..21341
/note="Alusg repeat: matches 5..304 of consensus"
21820..21882
/note="ORSL repeat: matches 369..427 of consensus"
21883..22200
/note="Aluub repeat: matches 1..312 of consensus"
22416..22682
/note="L2 repeat: matches 1572..1842 of consensus"
23587..23754
/note="MIR repeat: matches 3..172 of consensus"
24665..25041
/note="MIR11 repeat: matches 109..513 of consensus"
25058..25435
/note="MIR11 repeat: matches 101..471 of consensus"
25452..25623
/note="L2 repeat: matches 2113..2285 of consensus"
25658..25932
/note="L2 repeat: matches 2424..2731 of consensus"
26011..26079
/note="TIGGER2 repeat: matches 2633..2707 of consensus"
26377..26678
/note="AluY repeat: matches 1..302 of consensus"
27536..27844
/note="Aluud repeat: matches 1..302 of consensus"
29258..29705
/note="MIR1C repeat: matches 1..466 of consensus"
30014..30311
/note="Alusg repeat: matches 1..298 of consensus"
32010..32165
/note="LIPB3 repeat: matches 5901..6024 of consensus"
32166..32468
/note="Alusx repeat: matches 1..303 of consensus"
32469..32591
/note="LIPB3 repeat: matches 6024..6150 of consensus"
33112..33418
/note="AlusC repeat: matches 1..306 of consensus"
34819..34910
/note="MIR repeat: matches 41..131 of consensus"
35691..36075
/note="7 copies 55 mer 64% conserved"
35692..36075
/note="192 copies 2 mer aa 60% conserved"
35914..36033
/note="30 copies 4 mer aaag 75% conserved"
36279..36342
/note="16 copies 4 mer agag 76% conserved"
37040..37216
/note="LIPB16 repeat: matches 5308..5482 of consensus"
37217..37510
/note="Alusg repeat: matches 1..294 of consensus"
37511..38166
/note="LIPB16 repeat: matches 5482..6146 of consensus"
38996..39207
/note="MIR repeat: matches 46..251 of consensus"
39330..39714
/note="MIR1B repeat: matches 1..390 of consensus"
40020..40075
/note="14 copies 4 mer tcta 71% conserved"
41109..41437
/note="MIR11 repeat: matches 66..547 of consensus"
42960..43107
/note="L2 repeat: matches 2605..2748 of consensus"
43297..43583
/note="Alusg repeat: matches 11..297 of consensus"
43628..43713
/note="MIR repeat: matches 158..262 of consensus"
43714..44269
/note="MIR1F repeat: matches 1..541 of consensus"
44270..44325
/note="MIR repeat: matches 108..158 of consensus"
44618..44680
/note="MIR2D repeat: matches 489..552 of consensus"
44719..45218
/note="MIR2D repeat: matches 1..533 of consensus"
45344..45371
/note="14 copies 2 mer tg 96% conserved"
```


REFERENCE 5 (bases 1 to 166774)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Oct 22, 1999 this sequence version replaced gi:3450912.

Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplense@wustl.wustl.edu

Summary Statistics
Center project name: H_DJ0414A15

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
This clone from chromosome 14 was provided by Dr. Pieter de Jong, Roswell Park Cancer Institute, Human Genetics Department, Elm and Carlton Streets, Buffalo NY 14263-0001 USA.

SOURCE INFORMATION:
This clone was derived from human PAC library RPT-3, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
Base positions 27488 to 27520 of RP3-414A15 is represented by a single chemistry.

The clone sequenced to the left is RP4-693M11, 200 bp overlap; the clone sequenced to the right is RP1-240K6, 200 bp overlap. Actual start of this clone is at base position 197 of RP3-414A15; actual end is at base position 16578 of RP3-414A15.

FEATURES

source

1.166774
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q24.3"
/clone="RP3-414A15"
/clone_lib="RPT-3"
139..428
/rpt_family="Alu"
197..868
/db_xref="GI:3168733"
579..607
/rpt_family="AT_Rich"
593..607
/note="similar to EST AA652017 (NID:q2583669) ns47n02.s1"
709..802
/rpt_family="U6"
907..1206
/rpt_family="Alu"
1793..2473
/rpt_family="L1"

repeat_region 2474..2773
/rpt_family="Alu"
repeat_region 2774..2857
/rpt_family="L1"
repeat_region 2891..3148
/rpt_family="Alu"
repeat_region 3285..3403
/rpt_family="Alu"
repeat_region 3412..3455
/rpt_family="AT_Rich"
repeat_region 3552..3595
/rpt_family="L2"
repeat_region 3643..3933
/rpt_family="Alu"
repeat_region 4298..4334
/rpt_family="U2"
repeat_region 4473..4493
/rpt_family="AT_Rich"
repeat_region 4722..5036
/rpt_family="Alu"
repeat_region 5124..5159
/rpt_family="AT_Rich"
repeat_region 5161..5234
/rpt_family="L1"
repeat_region 5170..5400
/rpt_family="L1"
repeat_region 5526..5559
/rpt_family="(A)n"
repeat_region 5562..5606
/rpt_family="(TATC)n"
repeat_region 5607..5891
/rpt_family="Alu"
repeat_region 6147..6224
/rpt_family="MER1_type"
repeat_region 6423..6515
/rpt_family="MER1_type"
repeat_region 6626..6710
/rpt_family="MER4_group"
repeat_region 6711..7001
/rpt_family="Alu"
repeat_region 7002..7028
/rpt_family="(TTAAA)n"
repeat_region 7029..7423
/rpt_family="MER4_group"
repeat_region 7494..7791
/rpt_family="L1"
repeat_region 7793..8092
/rpt_family="Alu"
repeat_region 8100..8332
/rpt_family="Alu"
repeat_region 8372..8411
/rpt_family="L1"
repeat_region 8412..8710
/rpt_family="Alu"
repeat_region 8711..9087
/rpt_family="L1"
repeat_region 9088..9219
/rpt_family="Alu"
repeat_region 9220..9427
/rpt_family="L1"
repeat_region 9428..9560
/rpt_family="Alu"
repeat_region 9561..9592
/rpt_family="L1"
repeat_region 9593..9726
/rpt_family="Alu"
repeat_region 9727..9806
/rpt_family="L1"
repeat_region 9809..9948
/rpt_family="L2"
repeat_region 9949..10079
/rpt_family="MER2_type"
repeat_region 10080..10365

Query Match	7.9%	Score 39.8;	DB 85;	Length 166774;
Best Local Similarity	52.0%;	Pred. No. 1.4;		
Matches 89; Conservative	0;	Mismatches 82;	Indels 0;	Gaps 0;

RESULT	15				
AC009843					
LOCUS					
DEFINITION					
AC009843	127091 bp	DNA	HTG	28-FEB-2000	
Drosophila melanogaster chromosome 3 clone BACR01611 (D1045)					
RC11-98 01.G11 map 80D1-80D2 strain Y; cn bw sp, *** SEQUENCING IN					
PROGRESS ***; 93 unordered pieces.					

REFERENCE
AUTHORS

Batheoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C. M., Farfan, D. E., Galle, R., George, R. A., Harris, N. L.,
 Hinkle, A., Hoskins, R. A., Houston, K. A., Hummasti, S. R., Karra, K.,
 Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A. R.,
 Moshrefi, M., Nixon, K., Paclob, J. M., Park, S., Pfeiffer, B.,
 Richards, S., Sethi, H., Syirikas, R. R., Wan, K. H., Webster, D.,
 Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G. M.
 Sequencing of *Drosophila melanogaster*
 Title
 Journal
 Reference
 2 (bases 1 to 127091)

AUTHORS
Celnicker, S. E., Agapayni, A., Arcina, T. T., Baxter, E., Blazej, R. G.,
Butenheff, C., Champagne, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C. M., Farfan, D. E., Galle, R., George, R. A., Harris, N. L.,
Hoskins, R. A., Houston, K. A., Hummasti, S. R., Karia, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomontin, M. A., Mazda, P.,

TITLE	JOURNAL	COMMENT
1. The Role of the State in Economic Development	Journal of Economic Surveys	See page 100
2. The Impact of Globalization on the Environment	Environmental Economics and Organization	See page 100
3. The Effect of Trade Liberalization on Income Distribution	Journal of Development Economics	See page 100
4. The Role of the State in the Provision of Social Services	Journal of Public Economics	See page 100
5. The Impact of Financial Globalization on the Real Economy	Journal of International Money and Finance	See page 100
6. The Effect of Trade Liberalization on the Environment	Journal of Environmental Economics and Organization	See page 100
7. The Role of the State in the Provision of Health Services	Journal of Health Economics	See page 100
8. The Impact of Financial Globalization on the Financial System	Journal of International Money and Finance	See page 100
9. The Effect of Trade Liberalization on the Labor Market	Journal of Development Economics	See page 100
10. The Role of the State in the Provision of Education Services	Journal of Public Economics	See page 100

Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacelet, J.M., Park, S., Pfeiffer, B., Poon, L., Sequelra, A., Sethi, H., Snir, E., Swirski, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. Rubin, G.M.

On Feb 28, 2000 this sequence version replaced g1:6980150. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to buggerfruitfly@berkeley.edu. All contigs in this submission meet the following cutoffs: length >= 200 bases.

```

** NOTE: This is a 'working draft' sequence. It currently
** consists of 93 contigs. The true order of the pieces
** is not known and their order in this sequence record is
** arbitrary. Gaps between the contigs are represented as
** runs of N, but the exact sizes of the gaps are unknown.
** This record will be updated with the finished sequence
** as soon as it is available and the accession number will
** be preserved.

```

*	1	562:	contig of 562 bp in length
*	563	642:	gap of unknown length
*	643	1276:	contig of 634 bp in length
*	1277	1356:	gap of unknown length
*	1357	1983:	contig of 627 bp in length
*	1984	2053:	gap of unknown length
*	2054	2313:	contig of 250 bp in length
*	2314	2393:	gap of unknown length
*	2394	2865:	contig of 472 bp in length
*	2866	2945:	gap of unknown length
*	2946	3439:	contig of 548 bp in length
*	3494	3573:	gap of unknown length
*	3574	4071:	contig of 498 bp in length
*	4072	4151:	gap of unknown length
*	4152	4897:	contig of 746 bp in length
*	4898	4977:	gap of unknown length
*	4878	5419:	contig of 442 bp in length
*	5420	5499:	gap of unknown length
*	5500	6411:	contig of 972 bp in length
*	6472	6551:	gap of unknown length
*	6552	7505:	contig of 954 bp in length
*	7506	7585:	gap of unknown length
*	7586	8204:	contig of 619 bp in length
*	8205	8284:	gap of unknown length
*	8285	9209:	contig of 925 bp in length
*	9210	9289:	gap of unknown length
*	9290	9981:	contig of 692 bp in length
*	9982	10061:	gap of unknown length
*	10062	11200:	contig of 1139 bp in length
*	11201	11280:	gap of unknown length
*	11281	11987:	contig of 707 bp in length
*	11988	12067:	gap of unknown length
*	12068	12786:	contig of 719 bp in length
*	12787	12866:	gap of unknown length
*	12867	13456:	contig of 590 bp in length
*	13457	13536:	gap of unknown length
*	13537	14333:	contig of 797 bp in length
*	14334	14413:	gap of unknown length
*	14414	15065:	contig of 652 bp in length
*	15066	15145:	gap of unknown length
*	15146	15574:	contig of 429 bp in length
*	15575	15654:	gap of unknown length
*	15655	16170:	contig of 516 bp in length
*	16171	16230:	gap of unknown length
*	16231	16868:	contig of 618 bp in length
*	16869	16948:	gap of unknown length
*	16949	17712:	contig of 764 bp in length
*	17713	17792:	gap of unknown length
*	17793	19275:	contig of 1483 bp in length
*	19276	19355:	gap of unknown length
*	19356	20073:	contig of 718 bp in length
*	20074	20153:	gap of unknown length

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 16, 2001, 19:04:36 ; Search time 1566.45 Seconds
(without alignments)
197.488 Million cell updates/sec

Title: US-09-675-650-4
Perfect score: 20
Sequence: 1 gagtaggaagatagaacg 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	20	100.0	5435	88	AF103908 Homo sapi
c 2	20	100.0	164371	80	AL390239 Homo sapi
c 3	20	100.0	173831	80	AL359314 Homo sapi
c 4	20	100.0	267581	80	AL358573 Homo sapi
c 5	17.4	87.0	130244	63	AC013518 Homo sapi
c 6	17.4	87.0	130336	92	HS17K7 Human DNA
c 7	17.4	87.0	166525	88	AC022413 Homo sapi
c 8	17.4	87.0	176040	70	AC027548 Homo sapi

C	9	17.4	87.0	184494	77	AC087189	AC087189 Homo sapi
C	10	17.4	87.0	188433	81	AL450391	AL450391 Mus muscu
C	11	17.4	87.0	196840	67	AC023255	AC023255 Homo sapi
C	12	17.4	87.0	197663	87	AC009245	AC009245 Human Chr
C	13	17.4	87.0	210859	81	AL513347	AL513347 Mus muscu
C	14	17.4	85.0	191549	79	AL158210	AL158210 Homo sapi
C	15	16.8	84.0	792	9	A96152	A96152 Sequence 18
C	16	16.8	84.0	107271	14	AT78M16	AL390921 Arabidops
C	17	16.8	84.0	125077	84	HS162E17	AL160492 Homo sapi
C	18	16.8	84.0	152036	79	AL161776	AL161776 Homo sapi
C	19	16.8	84.0	153783	90	AL157937	AL157937 Human DNA
C	20	16.8	84.0	156002	71	AC036110	AC036110 Homo sapi
C	21	16.8	84.0	170720	87	AC009309	AC009309 Homo sapi
C	22	16.8	84.0	185573	62	AC010994	AC010994 Drosophi1
C	23	16.8	84.0	192104	61	AC011141	AC011141 Homo sapi
C	24	16.8	84.0	192199	68	AC021810	AC021810 Homo sapi
C	25	16.8	84.0	196470	68	AC024382	AC024382 Homo sapi
C	26	16.8	84.0	199905	79	AL354709	AL354709 Homo sapi
C	27	16.8	84.0	200573	66	AC021629	AC021629 Mus muscu
C	28	16.8	84.0	209285	63	AC013517	AC013517 Homo sapi
C	29	16.8	84.0	213112	81	AL513543	AL513543 Homo sapi
C	30	16.8	84.0	214894	69	AC025751	AC025751 Mus muscu
C	31	16.8	84.0	231464	89	AF111168	AF111168 Homo sapi
C	32	16.4	82.0	1018	72	AC054642	AC054642 Giardia 1
C	33	16.4	82.0	1037	71	AC034447	AC034447 Giardia 1
C	34	16.4	82.0	2911	94	AF155511	AF155511 Mus muscu
C	35	16.4	82.0	56163	87	AC010126	AC010126 Homo sapi
C	36	16.4	82.0	83511	12	AB013389	AB013389 Arabidops
C	37	16.4	82.0	89292	91	AP001252	AP001252 Homo sapi
C	38	16.4	82.0	100061	64	AC016551	AC016551 Homo sapi
C	39	16.4	82.0	107172	92	HS737M10	AL031388 Human DNA
C	40	16.4	82.0	136570	64	AC016555	AC016555 Homo sapi
C	41	16.4	82.0	143811	63	AC013617	AC013617 Homo sapi
C	42	16.4	82.0	146596	86	AC007631	AC007631 Genomic s
C	43	16.4	82.0	155313	62	AC011091	AC011091 Homo sapi
C	44	16.4	82.0	155342	77	AC084269	AC084269 Rattus no
C	45	16.4	82.0	155827	84	CNS05TCT	AL355099 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AF103908 5435 bp DNA PRI 14-AUG-2000
DEFINITION Homo sapiens non-coding RNA DD3 gene, exons 2, 3, and 4.
ACCESSION AF103908
VERSION AF103908.1 GI:6165974
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 5435)
AUTHORS Bussemakers,M.J., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaus,H.F., Schalken,J.A., Debruyne,F.M., Ru,N. and Isaacs,W.B.
DD3: a new prostate-specific gene, highly overexpressed in prostate
cancer
JOURNAL Cancer Res. 59 (23), 5975-5979 (1999)
MEDLINE 20072360
PUBMED 10606244
2 (bases 1 to 5435)
AUTHORS Bussemakers,M.J.G., Van Bokhoven,A., Verhaegh,G.W., Smit,F.P.,
Karthaus,H.F.M., Schalken,J.A., Debruyne,F.M.J., Ru,N. and
Isaacs,W.B.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1998) Urology Research Laboratory, University
Hospital Nijmegen, P.O. Box 9101, Nijmegen 6500 HB, Netherlands
FEATURES
SOURCE 1..5435
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"

repeat_region /map="9q21-q22"
1..78
/rpl_family="AluY"
/rpl_type-dispersed
exon 533..697
/note="alternative exon present in 5% of cDNA clones"
/number=2
repeat_region 1035..1294
/rpl_family="Alu"
/rpl_type-dispersed
join(1571..1753,1981..5435)
/note="transcript III"
mRNA /product="non-coding RNA DD3"
join(1571..1753,1981..2517)
/note="transcript I"
mRNA /product="non-coding RNA DD3"
join(1571..1753,1981..3579)
/note="transcript (major) II"
/product="non-coding RNA DD3"
1571..1753
exon /number=3
/number=4
exon 1981..5435
/number=4
polyA_signal 2495..2499
/note="transcript I"
polyA_site 2517
/note="transcript I"
polyA_signal 3533..3558
/note="transcript (major) II"
polyA_site 3579
/note="transcript (major) II"
repeat_region 5268..5423
/note="LINE"
/rpl_family="L1"
/rpl_type-dispersed
polyA_site 5435
/note="transcript III"
BASE COUNT 1546 a 1188 c 1036 g 1654 t 11 others
ORIGIN

Query Match 100.0%; Score 20; DB 88; Length 5435;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gagtaggaaggatagaacg 20
|||||
Db 1838 GAGTAGGAGGATGAGAAACG 1819

RESULT 2
LOCUS AL390239 164371 bp DNA HMG 09-MAR-2001
DEFINITION Homo sapiens chromosome 9 clone RP11-58J3, *** SEQUENCING IN
PROGRESS ***, 4 unordered pieces.
ACCESSION AL390239
VERSION AL390239.11 GI:13274794
KEYWORDS HMG; HMG_PHASE1; HMG_DRAFT; HMG_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 164371)
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 12, 2001 this sequence version replaced g1:13273805.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba58j3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 16337 bases at least Q40
Consensus quality: 163481 bases at least Q30
Consensus quality: 163577 bases at least Q20
Insert size: 164071; sum-of-contigs
Insert size: 166918; 4.2% error; agarose-fp
Quality coverage: 8.78x in Q20 bases; sum-of-contigs Quality
coverage: 9.07x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 20240: contig of 20240 bp in length
* 20241 20340: gap of 100 bp
* 20341 116100: contig of 95760 bp in length
* 116101 116200: gap of 100 bp
* 116201 118243: contig of 2043 bp in length
* 118244 118343: gap of 100 bp
* 118344 164371: contig of 46028 bp in length.
Location/Qualifiers
1. 164371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-58J3"
/clone_1lb="RP11-11.1"
1. 20240
/note="assembly_fragment:00838
fragment_chain:1
vector_end:SP6
vector_side:left"
20341. 116100
/note="assembly_fragment:01998
fragment_chain:1"
116201. 118243
/note="assembly_fragment:00916
fragment_chain:1"
118344. 164371
/note="assembly_fragment:00334
fragment_chain:1
clone_end:T7
vector_side:right"
BASE COUNT 50254 a 34905 c 32663 g 46249 t 300 others
ORIGIN
Query Match 100.0%; Score 20; DB 80; Length 164371;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gagtaggaagatagaacg 20
|||||
Db 33177 GAGTAGGAAGATAGAACC 33158

RESULT 3
AL359314/c 173831 bp DNA HTG 08-APR-2001
LOCUS Homo sapiens chromosome 9 clone RP11-108L4, *** SEQUENCING IN
DEFINITION PROGRESS ***; 3 unordered pieces.
AL359314
ACCESSION AL359314.12 GI:13396560
VERSION

KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS 1 (bases 1 to 173831)
TITLE Direct Submission
JOURNAL Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Mar 20, 2001 this sequence version replaced gi:13277120.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba108l4
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 173545 bases at least Q40
Consensus quality: 173599 bases at least Q30
Consensus quality: 173621 bases at least Q20
Insert size: 173631; sum-of-contigs
Insert size: 172123; 10.0% error; agarose-fp
Quality coverage: 10.35x in Q20 bases; sum-of-contigs Quality
coverage: 10.50x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 50595: contig of 50595 bp in length
* 50596 50695: gap of 100 bp
* 50696 157636: contig of 106941 bp in length
* 157637 157736: gap of 100 bp
* 157737 173831: contig of 16095 bp in length.
Location/Qualifiers
1. 173831
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-108L4"
/clone_1lb="RP11-11.1"
1. 50595
/note="assembly_fragment:04137
fragment_chain:1"
50696. 157636
/note="assembly_fragment:04236
fragment_chain:1"
157737. 173831
/note="assembly_fragment:01656"
BASE COUNT 52107 a 36469 c 35018 g 50037 t 200 others
ORIGIN
Query Match 100.0%; Score 20; DB 80; Length 173831;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gagtaggaagatagaacg 20
|||||
Db 156098 GAGTAGGAAGATAGAACC 156079

RESULT 4

AL358573	AL358573	267581 bp	DNA	HTG	15-APR-2001
LOCUS	Homo sapiens chromosome 9 clone RP11-133022, *** SEQUENCING IN PROGRESS ***	37 unordered pieces.			
ACCESSION	AL358573				
VERSION	AL358573.17	GI:13660951			
KEYWORDS	HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 267581)				
JOURNAL	McClay, K.				
COMMENT	Direct Submission Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Requests: clonerequests@sanger.ac.uk On Apr 17, 2001 this sequence version replaced gt:13398774. ----- Genome Center Center: Sanger Centre Center code: SC Web site: http://www.sanger.ac.uk Contact: humquery@sanger.ac.uk ----- Project Information Center project name: BA133022 ----- Summary Statistics Assembly program: XGAP; Version 4.5 Sequencing vector: plasmid; L08752; 100% of reads Chemistry: Dye-terminator Big Dye; 100% of reads Consensus quality: 256699 bases at least Q40 Consensus quality: 259744 bases at least Q30 Consensus quality: 261407 bases at least Q20 Insert size: 263981; sum-of-contigs Insert size: 135491; 19.3% error; agarose-fp Quality coverage: 5.51x in Q20 bases; sum-of-contigs Quality coverage: 12.23x in Q20 bases; agarose-fp ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 37 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * 1 25718: contig of 25718 bp in length * 25719 25818: gap of 100 bp * 25819 28567: contig of 3749 bp in length * 29568 29667: gap of 100 bp * 29668 40520: contig of 10853 bp in length * 40521 40620: gap of 100 bp * 40621 50002: contig of 9382 bp in length * 50003 50103: gap of 100 bp * 50103 50942: contig of 4840 bp in length * 50943 55042: gap of 100 bp * 55043 59067: contig of 4025 bp in length * 59068 59167: gap of 100 bp * 59168 72327: contig of 13160 bp in length * 72328 72427: gap of 100 bp * 72428 79396: contig of 6699 bp in length * 79397 79496: gap of 100 bp * 79497 82738: contig of 3242 bp in length * 82739 82838: gap of 100 bp * 82839 87543: contig of 4705 bp in length * 87544 87643: gap of 100 bp * 87644 89847: contig of 2204 bp in length * 89848 89947: gap of 100 bp * 89948 92376: contig of 2429 bp in length * 92377 92476: gap of 100 bp * 92477 99799: contig of 7323 bp in length * 99800 99899: gap of 100 bp * 99900 102117: contig of 2218 bp in length * 102118 102217: gap of 100 bp				

	*	1023218	109330:	contig of	7113 bp	in length
	*	109331	109430:	gap of	100 bp	
	*	109431	112187:	contig of	2757 bp	in length
	*	112188	112287:	gap of	100 bp	
	*	112288	116407:	contig of	4140 bp	in length
	*	116408	116507:	gap of	100 bp	
	*	116508	119386:	contig of	2879 bp	in length
	*	119387	119486:	gap of	100 bp	
	*	121870	121869:	contig of	2363 bp	in length
	*	121870	125546:	contig of	3577 bp	in length
	*	125547	125646:	gap of	100 bp	
	*	125647	128890:	contig of	3344 bp	in length
	*	128891	129090:	gap of	100 bp	
	*	129091	134055:	contig of	4965 bp	in length
	*	134056	134155:	gap of	100 bp	
	*	134156	138314:	contig of	4159 bp	in length
	*	138315	138414:	gap of	100 bp	
	*	138415	140612:	contig of	2198 bp	in length
	*	140613	140712:	gap of	100 bp	
	*	140713	143296:	contig of	2584 bp	in length
	*	143297	143396:	gap of	100 bp	
	*	143397	148113:	contig of	4717 bp	in length
	*	148114	148213:	gap of	100 bp	
	*	148214	151009:	contig of	2796 bp	in length
	*	151010	151109:	gap of	100 bp	
	*	151110	156616:	contig of	5507 bp	in length
	*	156617	156716:	gap of	100 bp	
	*	156717	158763:	contig of	2047 bp	in length
	*	158764	158863:	gap of	100 bp	
	*	158864	162129:	contig of	3266 bp	in length
	*	162130	162229:	gap of	100 bp	
	*	162230	164443:	contig of	2214 bp	in length
	*	164444	165453:	gap of	100 bp	
	*	165454	169800:	contig of	5257 bp	in length
	*	169801	169900:	gap of	100 bp	
	*	169901	251445:	contig of	81645 bp	in length
	*	251546	251645:	gap of	100 bp	
	*	251646	259029:	contig of	7384 bp	in length
	*	259030	259129:	gap of	100 bp	
	*	259130	262871:	contig of	3742 bp	in length
	*	262872	262971:	gap of	100 bp	
	*	262972	265209:	contig of	2238 bp	in length
	*	265210	265309:	gap of	100 bp	
	*	265310	267581:	contig of	2272 bp	in length.
FEATURES						
source						
			1.. 267581			
			/organism="Homo sapiens"			
			/db_xref="taxon:9606"			
			/chromosome="9"			
			/clone="RP11-133022"			
			/clone_11b="RPC1-11.1"			
			1.. 25718			
misc_feature			/note="assembly_fragment:00645			
			fragment_chain:1			
			clone_end:17			
			vector_side:left			
misc_feature			25819..29567			
			/note="assembly_fragment:01660			
			fragment_chain:1			
misc_feature			29668..40520			
			/note="assembly_fragment:03080			
			fragment_chain:1			
misc_feature			40621..50002			
			/note="assembly_fragment:00771			
			fragment_chain:1			
misc_feature			50103..54942			
			/note="assembly_fragment:02629			
			fragment_chain:1			
misc_feature			55043..59067			
			/note="assembly_fragment:04565			
			fragment_chain:1			
misc_feature			59168..72327			

```

/note="assembly-fragment:00223
fragment_chain:1"
misc_feature
72428..79396
/note="assembly-fragment:03318
fragment_chain:1"
misc_feature
79497..82738
/note="assembly-fragment:00224
fragment_chain:2"
misc_feature
82839..87543
/note="assembly-fragment:03050
fragment_chain:2"
misc_feature
87644..89847
/note="assembly-fragment:01948
fragment_chain:3"
misc_feature
89948..92376
/note="assembly-fragment:02191
fragment_chain:3"
misc_feature
92477..99799
/note="assembly-fragment:00010"
99900..102117
/note="assembly-fragment:00672"
102218..109330
/note="assembly-fragment:01326"
109431..112187
/note="assembly-fragment:01498"
112288..116407
/note="assembly-fragment:01607"
116508..119386
/note="assembly-fragment:01899"
119487..121869
/note="assembly-fragment:02068"
121970..125546
/note="assembly-fragment:02469"
125647..128990
/note="assembly-fragment:02502"
129091..134055
/note="assembly-fragment:02669"
134156..138314
/note="assembly-fragment:02757"
138415..140612
/note="assembly-fragment:02776"
140713..143296
/note="assembly-fragment:03127"
143397..148113
/note="assembly-fragment:03634"
148214..151009
/note="assembly-fragment:03898"
151110..156616
/note="assembly-fragment:04220"
156717..158763
/note="assembly-fragment:04276"
158864..162129
/note="assembly-fragment:04418"
162230..164443
/note="assembly-fragment:04428"

Query Match      100.0%; Score 20; DB 80; Length 267581;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 gagtaggaagatagaagac 20
         |||||||
Db 166039 GAGTAGGAGATAGAACG 166058

RESULT      5
LOCUS       AC013518      130244 bp      DNA
DEFINITION  Homo sapiens clone RP11-115N3, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC013518
VERSION     AC013518.2 GI:7533963
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.

```

```

ORGANISM      Homo sapiens
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS       1 (bases 1 to 130244)
              Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL       Homo sapiens chromosome, clone RP11-115N3
              Unpublished
              2 (bases 1 to 130244)
              Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
              Baldwin,J., Barua,N., Beckerly,R., Boguslavsky,I., Bouhgaier,B.,
              Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
              Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
              Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
              Galagan,J.C., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
              Howland,J.C., Johnson,R., Jones,C., Kann,J., Karats,A., Klein,J.,
              Lehoczkzy,J., Lien,C., Locke,K., McDonald,P., Margulis,N.,
              McEwan,P., McGuirk,A., McKernan,K., McLaughlin,J., Meldim,J.,
              Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
              Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
              Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
              Testaye,S., Titrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
              Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
              Direct Submission
              Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
              On Apr 11, 2000 this sequence version replaced gi:6403763.
              All repeats were identified using RepeatMasker:
              Smit, A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html
              Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
              Project Information
              Center project name: 115_N3
              Center clone name: 115_N3

NOTE: This record contains 151 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
1
729 828: gap of 100 bp
829 1582: contig of 754 bp in length
1583 1682: gap of 100 bp
1683 2482: contig of 800 bp in length
2483 2582: gap of 100 bp
2583 3469: contig of 787 bp in length
3470 4225: contig of 756 bp in length
4226 4325: gap of 100 bp
4326 5088: contig of 763 bp in length
5089 5188: gap of 100 bp
5189 5950: contig of 762 bp in length
5951 6050: gap of 100 bp
6051 6818: contig of 768 bp in length
6819 7686: contig of 768 bp in length
7687 7786: gap of 100 bp
7787 8564: contig of 778 bp in length
8565 8664: gap of 100 bp
8665 9409: contig of 745 bp in length
9410 9509: gap of 100 bp
9510 10263: contig of 754 bp in length
10264 10363: gap of 100 bp

```

```

* 10364 11149: contig of 786 bp in length
* 11150 11249: gap of 100 bp
* 11250 12019: contig of 770 bp in length
* 12020 12119: gap of 100 bp
* 12120 12860: contig of 741 bp in length
* 12861 12960: gap of 100 bp
* 12961 13712: contig of 752 bp in length
* 13713 13812: gap of 100 bp
* 13813 14574: contig of 762 bp in length
* 14575 14674: gap of 100 bp
* 14675 15443: contig of 769 bp in length
* 15444 15543: gap of 100 bp
* 15544 16289: contig of 746 bp in length
* 16290 16389: gap of 100 bp
* 16390 17170: contig of 781 bp in length
* 17171 17270: gap of 100 bp
* 17271 18032: contig of 762 bp in length
* 18033 18132: gap of 100 bp
* 18133 18903: contig of 771 bp in length
* 18904 19003: gap of 100 bp
* 19004 19795: contig of 792 bp in length
* 19796 19895: gap of 100 bp
* 19896 20654: contig of 759 bp in length
* 20655 20734: gap of 100 bp
* 20735 21498: contig of 744 bp in length
* 21499 21598: gap of 100 bp
* 21599 22366: contig of 768 bp in length
* 22367 22466: gap of 100 bp
* 22467 23232: contig of 766 bp in length
* 23233 23332: gap of 100 bp
* 23333 24101: contig of 769 bp in length
* 24102 24201: gap of 100 bp
* 24202 24911: contig of 710 bp in length
* 24912 25011: gap of 100 bp
* 25012 25793: contig of 782 bp in length
* 25794 25893: gap of 100 bp
* 25894 26674: contig of 781 bp in length
* 26675 26774: gap of 100 bp
* 26775 27524: contig of 750 bp in length
* 27525 27624: gap of 100 bp
* 27625 28401: contig of 777 bp in length
* 28402 28501: gap of 100 bp
* 28502 29257: contig of 756 bp in length
* 29258 29357: gap of 100 bp
* 29358 30136: contig of 779 bp in length
* 30137 30236: gap of 100 bp
* 30237 30998: contig of 762 bp in length
* 30999 31098: gap of 100 bp
* 31099 31858: contig of 760 bp in length
* 31859 31958: gap of 100 bp
* 31959 32741: contig of 783 bp in length
* 32742 32841: gap of 100 bp
* 32842 33598: contig of 758 bp in length
* 33600 33699: gap of 100 bp
* 33700 34480: contig of 781 bp in length
* 34481 34580: gap of 100 bp
* 34581 35344: contig of 764 bp in length
* 35345 35444: gap of 100 bp
* 35445 36202: contig of 758 bp in length
* 36203 36302: gap of 100 bp
* 36303 37026: contig of 724 bp in length
* 37027 37126: gap of 100 bp
* 37127 37910: contig of 784 bp in length
* 37911 38010: gap of 100 bp
* 38011 38772: contig of 762 bp in length
* 38773 38872: gap of 100 bp
* 38873 39647: contig of 775 bp in length
* 39648 39747: gap of 100 bp
* 39748 40514: contig of 767 bp in length
* 40515 40614: gap of 100 bp
* 40615 41379: contig of 765 bp in length
* 41380 41479: gap of 100 bp
* 41480 42236: contig of 757 bp in length

```

```

* 42237 42336: gap of 100 bp
* 42337 43080: contig of 744 bp in length
* 43081 43180: gap of 100 bp
* 43181 43962: contig of 782 bp in length
* 43963 44062: gap of 100 bp
* 44063 44837: contig of 775 bp in length
* 44838 44937: gap of 100 bp
* 44938 45723: contig of 786 bp in length
* 45724 45823: gap of 100 bp
* 45824 46607: contig of 784 bp in length
* 46608 46707: gap of 100 bp
* 46708 47493: contig of 786 bp in length
* 47494 47593: gap of 100 bp
* 47594 48327: contig of 734 bp in length
* 48328 48427: gap of 100 bp
* 48428 49171: contig of 744 bp in length
* 49172 49271: gap of 100 bp
* 49272 50018: contig of 747 bp in length
* 50019 50118: gap of 100 bp
* 50119 50886: contig of 768 bp in length
* 50887 50986: gap of 100 bp
* 50987 51755: contig of 769 bp in length
* 51756 51855: gap of 100 bp
* 51856 52633: contig of 778 bp in length
* 52634 52733: gap of 100 bp
* 52734 53508: contig of 775 bp in length
* 53509 53608: gap of 100 bp
* 53609 54371: contig of 763 bp in length
* 54372 54471: gap of 100 bp
* 54472 55232: contig of 761 bp in length
* 55233 55332: gap of 100 bp
* 55333 56121: contig of 789 bp in length
* 56122 56221: gap of 100 bp
* 56222 56972: contig of 751 bp in length
* 56973 57072: gap of 100 bp
* 57073 57848: contig of 776 bp in length
* 57849 57948: gap of 100 bp
* 57949 58698: contig of 750 bp in length
* 58699 58798: gap of 100 bp
* 58799 59568: contig of 770 bp in length
* 59569 59668: gap of 100 bp
* 59669 60436: contig of 768 bp in length
* 60437 60536: gap of 100 bp
* 60537 61302: contig of 766 bp in length
* 61303 61402: gap of 100 bp
* 61403 62167: contig of 765 bp in length
* 62168 62267: gap of 100 bp
* 62268 63043: contig of 782 bp in length

```

Query Match 87.0%; Score 17.4; DB 63; Length 130244;
 Best Local Similarity 94.7%; Pred. No. 1.9e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gactaggaagatagaac 19
 Db 34364 GAGTAGGAAGGAAGAAC 34346

RESULT 6
 HS17K7 130336 bp DNA PRI 07-FEB-2000
 LOCUS
 DEFINITION Human DNA sequence from clone RPI-17K7 on chromosome 11p13.
 Contains three novel genes, ESTs, STSS and GSSs, complete sequence.
 ACCESSION AL035078
 AL035078
 KEYWORDS
 VERSSION
 AL035078.32 GI:5805137
 HTG.
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 130336)
 AUTHORS Donnelly/S.
 TITLE Direct Submission

JOURNAL

COMMENT

Submitted (05-JAN-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Aug 30, 1999 this sequence version replaced g1:5804869.
 This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated by the Sanger Centre from a human chromosome 11 bacterial clone contig constructed by Niederfuhr, A. et al.
<http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=9790764&form=6&db=m&dbp=b>
 RPI-17K7 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://dacpac.med.buffalo.edu/>
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RPI-17K7. The true right end of clone RPI-65P5 is at 20899 in this sequence.
 Location/Qualifiers

FEATURES

Source

```

1..130336
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="p13"
/clone="RPI-17K7"
1..73
/note="AluSp/q repeat: matches 234. .306 of consensus"
3558..3754
/note="MER58A repeat: matches 1. .224 of consensus"
4589..4624
/note="18 copies 2 mer tt 88% conserved"
5389..5506
/note="MER91A repeat: matches 46. .170 of consensus"
7702..7889
/note="MER30 repeat: matches 1. .199 of consensus"
10794..10869
/note="19 copies 4 mer gata 68% conserved"
11304..11601
/note="AluSc repeat: matches 2. .299 of consensus"
11854..11906
/note="MLT2CA repeat: matches 455. .508 of consensus"
11909..11948
/note="10 copies 4 mer atcc 87% conserved"
11982..12017
/note="9 copies 4 mer catc 83% conserved"
12084..12455
/note="MLT2CA repeat: matches 1. .444 of consensus"
13610..13846
/note="L1PA15 repeat: matches 5653. .5901 of consensus"
14161..14434
/note="L1PA15 repeat: matches 5901. .6156 of consensus"
14517..14693
/note="match: GSS: Em:A0830841"
complement(14558..14988)
/note="match: GSS: Em:A0321591"
14999..15026
/note="7 copies 4 mer caca 92% conserved"
15027..15325
/note="L1ME3A repeat: matches 5813. .6112 of consensus"

```

```

repeat_region 18920..18997
/note="AluS repeat: matches 3. .80 of consensus"
repeat_region 20469..20856
/note="WSTD repeat: matches 1. .394 of consensus"
repeat_region 20889..20928
/note="20 copies 2 mer tg 80% conserved"
mRNA join(<1267..21380,21558..>21684)
/note="match: ESTs: Em:AA83698"
/note="match: ESTs: Em:AA905415"
/evidence="not_experimental"
/product="dJ17K7.1 (novel protein)"
join(21267..21380,21558..21684)
/gene="dJ17K7.1"
22693..22772
/note="20 copies 4 mer ttct 68% conserved"
repeat_region 23117..23220
/note="26 copies 4 mer cctt 91% conserved"
repeat_region 23911..24131
/note="MER58A repeat: matches 3. .224 of consensus"
25132..25463
/note="MER2 repeat: matches 1. .345 of consensus"
27272..27657
/note="match: GSS: Em:A0560386"
complement(join(27441..27646,28324..28356,29647..>29713))
/gene="dJ17K7.2"
/note="match: ESTs: Em:AA905415"
/evidence="not_experimental"
/product="dJ17K7.2 (novel protein)"
complement(27441..29713)
/gene="dJ17K7.2"
complement(27441)
/gene="dJ17K7.2"
complement(27453..27458)
/gene="dJ17K7.2"
complement(join(27622..27646,28324..28356,29647..>29713))
/gene="dJ17K7.2"
/note="partially supported by FGENES"
/codon_start=3
/evidence="not_experimental"
/product="dJ17K7.2 (novel protein)"
/protein_id="CAB72348.1"
/db_xref="GI:6957461"
/translational="ILYVNLSSVEQMGRSPGRAVHRKIVAKILLPREILEDKC"
27806..28292
/note="LRR40A repeat: matches 1. .501 of consensus"
28955..29060
/note="MER69 repeat: matches 1. .108 of consensus"
30874..31250
/note="match: GSS: Em:A032621"
31251..31418
/note="L1MA2 repeat: matches 6141. .6308 of consensus"
31531..31727
/note="L1PA7 repeat: matches 5949. .6145 of consensus"
join(<33244..33419,34150..34458)
/gene="dJ17K7.3"
/note="isoform 2"
match: ESTs: Em:A1027739 Em:AA725753 Em:AA921932
Em:AA926690 Em:AA905751"
/evidence="not_experimental"
/product="dJ17K7.3 (novel protein)"
join(<33244..33419,37531..37868)
/gene="dJ17K7.3"
/note="isoform 1"
match: ESTs: Em:AA927867 Em:AA904797 Em:AA301105
Em:AA926677 Em:AA883570 Em:AA904519 Em:A1026743
Em:AA812900 Em:AA910052 Em:AA962351"
/evidence="not_experimental"
/product="dJ17K7.3 (novel protein)"
33244..37868
/gene="dJ17K7.3"
33961..34197
/note="MLT1G repeat: matches 1. .251 of consensus"
34438..34443
polyA_signal

```

```

polyA_site      /gene="dj17k7.3"
                 /gene="dj17k7.3"
repeat_region    /note="19 copies 2 mer aa 81% conserved"
                 /note="8 copies 4 mer aaac 87% conserved"
mRNA             join(<35606..35675,37531..37868)
                 /gene="dj17k7.3"
                 /note="isoform 3"
                 match: ESTs: Em:A1215480"
                 /evidence=not_experimental
                 /product="dj17k7.3 (novel protein)"
repeat_region    /note="FRAM repeat: matches 1..176 of consensus"
                 37247..37422
repeat_region    /note="19 copies 2 mer aa 100% conserved"
                 37795..37832
polyA_signal     37846..37851
                 /gene="dj17k7.3"
polyA_site      37868
                 /gene="dj17k7.3"
repeat_region    /note="13 copies 4 mer tgtg 80% conserved"
                 37884..37935
repeat_region    /note="26 copies 2 mer gt 82% conserved"
                 37885..37936
repeat_region    /note="26 copies 2 mer gt 82% conserved"
                 38937..39020
repeat_region    /note="LIMB7 repeat: matches 5534..5618 of consensus"
                 39399..39919
repeat_region    /note="LIMB7 repeat: matches 5614..6158 of consensus"
                 41155..41284
repeat_region    /note="MER21B repeat: matches 12..135 of consensus"
                 /complement(43633..44023)
misc_feature     /note="match: STS: Em:Z67040"
                 43704..43761
repeat_region    /note="29 copies 2 mer tg 94% conserved"
                 43707..43762
repeat_region    /note="14 copies 4 mer gtgt 94% conserved"
                 /complement(44739..45379)
misc_feature     /note="match: GSS: Em:AQ319144"
                 47762..47783
repeat_region    /note="11 copies 2 mer ca 100% conserved"

Query Match      87.0% Score 17.4; DB 92; Length 130336;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gagtagaagatagaac 19
    |||||
Db 68214 GAGTTGAGAGATGAGAAC 68196

RESULT 7
AC022413/c      AC022413 166525 bp DNA PRI 25-JAN-2001
LOCUS           Homo sapiens chromosome 5 clone CTC-205M6, complete sequence.
DEFINITION      AC022413
ACCESSION       AC022413.4 GI:12484317
VERSION         HTG.
KEYWORDS        human.
SOURCE          Homo sapiens
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE       1 (bases 1 to 166525)
AUTHORS         DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE           Direct Submission
JOURNAL         Unpublished
AUTHORS         2 (bases 1 to 166525)
REFERENCE       DOE Joint Genome Institute.
TITLE           Direct Submission
JOURNAL         Submitted (03-FEB-2000) Production Sequencing Facility, DOE Joint
                 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS         3 (bases 1 to 166525)
REFERENCE       DOE Joint Genome Institute and Stanford Human Genome Center.

```

```

TITLE          Direct Submission
JOURNAL        Submitted (25-JAN-2001) DOE Joint Genome Institute, 2800 Mitchell
COMMENT        Drive, Walnut Creek, CA 94598, USA
               On Jan 25, 2001 this sequence version replaced gl:7711721.
               Draft Sequence Produced by DOE Joint Genome Institute
               www.jgi.doe.gov
               Finishing Completed at Stanford Human Genome Center
               www-shgc.stanford.edu
               Quality: Phrap Quality >=40 99.7% of Sequence:
               Estimated Total Number of Errors is 0.4.
               STS Content:
               WI-15997 G21714.

FEATURES
  source
    1..166525
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="5"
    /clone="CRC-205M6"

BASE COUNT      47370 a 36137 c 37038 g 45980 t
ORIGIN

Query Match      87.0% Score 17.4; DB 88; Length 166525;
Best Local Similarity 94.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gagtagaagatagaac 19
    |||||
Db 115493 GAGTAGAGAGATGAGAAC 115475

RESULT 8
AC027548/c      AC027548 176040 bp DNA HTG 16-MAR-2001
LOCUS           Homo sapiens chromosome 4 clone RP11-23K2 map 4, WORKING DRAFT
DEFINITION      AC027548
ACCESSION       AC027548.3 GI:13357512
VERSION         HTG; HTGS; PHASE1; HTGS_DRAFT.
KEYWORDS        human.
SOURCE          Homo sapiens
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE       1 (bases 1 to 176040)
AUTHORS         Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE           Homo sapiens chromosome 4, clone RP11-23K2
JOURNAL        Unpublished
AUTHORS         2 (bases 1 to 176040)
REFERENCE       Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
                 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
                 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
                 Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
                 Collamore,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S.,
                 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
                 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
                 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
                 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
                 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
                 Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
                 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeeters,R.,
                 Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mleaga,V., Morrow,J.,
                 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
                 O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
                 Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
                 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
                 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
                 Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
                 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
                 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE           Direct Submission
JOURNAL        Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome
                 Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT        On Mar 16, 2001 this sequence version replaced gl:7658429.

```

All repeats were identified using RepeatMasker:

Smiley, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8919

Center clone name: 23_K-2

----- Summary Statistics

Sequencing vector: M13; M77815; 92% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 16542 bases at least Q40

Consensus quality: 171629 bases at least Q30

Consensus quality: 173620 bases at least Q20

Insert size: 171000; agarose-fp

Insert size: 174740; sum-of-contigs

Quality coverage: 4.8 in Q20 bases; agarose-fp

Quality coverage: 4.7 in Q20 base.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 14 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 9984: contig of 9984 bp in length
9985 10084: gap of 100 bp
10085 11360: contig of 1276 bp in length
11361 11460: gap of 100 bp
11461 12753: contig of 1293 bp in length
12754 12853: gap of 100 bp
12854 14760: contig of 1907 bp in length
14761 14860: gap of 100 bp
14861 17146: contig of 2286 bp in length
17147 17246: gap of 100 bp
17247 20168: contig of 2922 bp in length
20169 20268: gap of 100 bp
20269 23632: contig of 3364 bp in length
23633 23732: gap of 100 bp
23733 30148: contig of 6416 bp in length
30149 30248: gap of 100 bp
30249 39448: contig of 9200 bp in length
39449 39548: gap of 100 bp
39549 49092: contig of 9544 bp in length
49093 49192: gap of 100 bp
49193 62340: contig of 13148 bp in length
62341 62440: gap of 100 bp
62441 77663: contig of 15223 bp in length
77664 77763: gap of 100 bp
77764 93527: contig of 15764 bp in length
93528 93627: gap of 100 bp
93628 176040: contig of 82413 bp in length.

Location/Qualifiers

1. 176040

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="RP11-23K2"

/clone_lib="RPCI-11 Human Male BAC"

1. 3984

/note="assembly_fragment"

clone_end:SP6

vector_side:left"

10085. 11360

/note="assembly_fragment"

11461. 12753

misc_feature

misc_feature

misc_feature

misc_feature /note="assembly_fragment"

12854. 14760

misc_feature /note="assembly_fragment"

14861. 17146

misc_feature /note="assembly_fragment"

17247. 20168

misc_feature /note="assembly_fragment"

20269. 23632

misc_feature /note="assembly_fragment"

23733. 30148

misc_feature /note="assembly_fragment"

30249. 39448

misc_feature /note="assembly_fragment"

39549. 49092

misc_feature /note="assembly_fragment"

49193. 62340

misc_feature /note="assembly_fragment"

62441. 77663

misc_feature /note="assembly_fragment"

77764. 93527

misc_feature /note="assembly_fragment"

93628. 176040

misc_feature /note="assembly_fragment"

/note="assembly_fragment"

BASE COUNT 49032 a 37924 c 37537 g 50226 t 1321 others

ORIGIN

Query Match 87.0%; Score 17.4; DB 70; Length 176040;

Best Local Similarity 94.7%; Pred. No. 1.9e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gagtaggaagatagaac 19
|||||
Db 122514 GAGTGGAGGATGAAAC 122496

RESULT 9
AC087189/c

LOCUS AC087189

DEFINITION Homo sapiens chromosome 16 clone RP11-43J5, WORKING DRAFT SEQUENCE,

26 unordered pieces.

ACCESSION AC087189

VERSION AC087189.1 GI:11693349

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 184494)

AUTHORS DOE Joint Genome Institute.

TITLE Unpublished

JOURNAL 2 (bases 1 to 184494)

REFERENCE DOE Joint Genome Institute.

AUTHORS Direct Submission

TITLE Submitted (13-DEC-2000) Production Sequencing Facility, DOE Joint

JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT -----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: http://www.jgi.doe.gov

Project Information

Center Project Name: 0

Center clone name: RPCI-11_43L5

Summary Statistics

Consensus quality: 132143 bases at least Q40

Consensus quality: 152435 bases at least Q30

Consensus quality: 161165 bases at least Q20

Estimated insert size: 183300; agarose-fp estimation

Estimated insert size: 181994; sum-of-contigs estimation

Quality coverage: 3.79 in Q20 bases; agarose-fp estimation

Quality coverage: 3.82 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently consists of 26 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
* 1 1270: contig of 1270 bp in length
* 1370: gap of unknown length
* 1371 1370: gap of unknown length
* 1371 2456: contig of 1086 bp in length
* 2457 2536: gap of unknown length
* 2537 4282: contig of 1726 bp in length
* 4283 4383: gap of unknown length
* 4383 5945: contig of 1563 bp in length
* 5946 6046: gap of unknown length
* 6046 7845: contig of 1800 bp in length
* 7846 7945: gap of unknown length
* 7946 9439: contig of 1494 bp in length
* 9440 9539: gap of unknown length
* 9540 11087: contig of 1548 bp in length
* 11088 11187: gap of unknown length
* 11188 12620: contig of 1433 bp in length
* 12621 12720: gap of unknown length
* 12721 14195: contig of 1475 bp in length
* 14196 14295: gap of unknown length
* 14296 16713: contig of 2418 bp in length
* 16714 16813: gap of unknown length
* 16814 19718: contig of 2805 bp in length
* 19719 23668: contig of 3950 bp in length
* 23669 26246: gap of unknown length
* 26247 26347: gap of unknown length
* 26348 30889: contig of 4543 bp in length
* 30890 30989: gap of unknown length
* 30990 36269: contig of 5280 bp in length
* 36270 36369: gap of unknown length
* 36370 38426: contig of 2057 bp in length
* 38427 38526: gap of unknown length
* 38527 42742: contig of 4216 bp in length
* 42743 42843: gap of unknown length
* 42844 50186: contig of 7344 bp in length
* 50187 50286: gap of unknown length
* 50287 58054: contig of 7768 bp in length
* 58055 58155: gap of unknown length
* 58156 66543: contig of 8389 bp in length
* 66544 66644: gap of unknown length
* 66645 77179: contig of 10536 bp in length
* 77180 77279: gap of unknown length
* 77280 91273: contig of 13994 bp in length
* 91274 91373: gap of unknown length
* 91374 106469: contig of 15096 bp in length
* 106470 106569: gap of unknown length
* 106570 120018: contig of 13449 bp in length
* 120019 120118: gap of unknown length
* 120119 137271: contig of 17153 bp in length
* 137272 137372: gap of unknown length
* 137373 184494: contig of 47123 bp in length.
```

FEATURES

Source

```
1. .184494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-43L5"
/collection="RP11-43L5"
BASE COUNT 53403 a 38377 c 39440 g 50374 t 2900 others
ORIGIN
```

Query Match 87.0%; Score 17.4; DB 77; Length 184494;
Best Local Similarity 94.7%; Pred. No. 2e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 gagtagaagatagaac 19
|||
Db 93481 GATTAGGAGATAGAAC 93463

RESULT 10
AL450391/c DNA HTG 09-MAR-2001
LOCUS
DEFINITION Mus musculus chromosome X clone RP23-38B5, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
ACCESSION AL450391.3 GI:13121461
VERSION AL450391.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (03-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Feb 23, 2001 this sequence version replaced gi:12581064.
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouseghar.mrc.ac.uk
----- Project Information
Center project name: DM38B5
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: Plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 182819 bases at least Q40
Consensus quality: 184615 bases at least Q30
Insert size: 186433; sum-of-contigs
Insert size: 178321; 8.6% error; agarose-IP
Quality coverage: 6.95% in Q20 bases; sum-of-contigs Quality
coverage: 7.45% in Q20 bases; agarose-IP

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

Source

```
1. .184433
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-38B5"
/clone_lib="RP11-23"
misc_feature 1. .3839
/note="assembly_fragment:01310
fragment_chain:1"
misc_feature 3940. .7493
/note="assembly_fragment:02284
fragment_chain:1"
misc_feature 7594. .12260
/note="assembly_fragment:00930
fragment_chain:1"
misc_feature 12361. .15763
/note="assembly_fragment:00752
fragment_chain:1"
misc_feature 15864. .24683
/note="assembly_fragment:02693
fragment_chain:1"
misc_feature 24784. .36625
/note="assembly_fragment:02604
```

```

misc_feature      fragment_chain:1"
                    36726..40555
                    /note="assembly_fragment:00724
                    fragment_chain:2"
misc_feature      40656..53507
                    /note="assembly_fragment:00525
                    fragment_chain:2"
misc_feature      53608..62434
                    /note="assembly_fragment:00844
                    fragment_chain:2"
misc_feature      62535..75598
                    /note="assembly_fragment:00153
                    fragment_chain:2"
misc_feature      76089..82344
                    /note="assembly_fragment:02835
                    fragment_chain:2"
misc_feature      82445..128364
                    /note="assembly_fragment:01348
                    fragment_chain:3"
misc_feature      128465..131120
                    /note="assembly_fragment:00806
                    fragment_chain:3"
misc_feature      131221..134024
                    /note="assembly_fragment:03427
                    fragment_chain:3"
misc_feature      134125..140441
                    /note="assembly_fragment:02048
                    fragment_chain:3"
misc_feature      140542..152665
                    /note="assembly_fragment:02156
                    fragment_chain:3"
misc_feature      152766..163506
                    /note="assembly_fragment:03380
                    fragment_chain:4"
misc_feature      163607..167405
                    /note="assembly_fragment:01016
                    fragment_chain:4"
misc_feature      167506..176298
                    /note="assembly_fragment:00910
                    fragment_chain:4"
misc_feature      176399..178490
                    /note="assembly_fragment:02838
                    fragment_chain:4"
misc_feature      178591..188433
                    /note="assembly_fragment:02018
                    clone_end:sp6
                    vector_side:right"
BASE COUNT      54216 a 38568 c 39683 g 53957 t 2009 others
ORIGIN
Query Match      87.0%; Score 17.4; DB 81; Length 188433;
Best Local Similarity 94.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gagtaggaagatagaac 19
Db 148542 GAGAAGAGAGATAGAAAC 148524

```

```

RESULT 11
AC023255
LOCUS      AC023255 196840 bp DNA HTG 25-JUN-2000
DEFINITION Homo sapiens clone Rpl1-754A4, WORKING DRAFT SEQUENCE, 33 unordered
            pieces.
ACCESSION  AC023255 GI:8705097
VERSION    AC023255.3
KEYWORDS   HTG, HTGS_PHASE1, HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 196840)

```

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone Rpl1-754A4
Unpublished
2 (bases 1 to 196840)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burrell,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearlano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kam,L., Karatas,A., Klein,J.,
Lander,T., Lechoczky,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Mcdonald,P., Margulis,N., McEwen,P., McGurk,A., McKernan,K.,
McPeeters,R., Melidrim,J., Menes,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Ollivar,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 25, 2000 this sequence version replaced gi:7534004.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 754_A4
Center clone name: 16561
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 170430 bases at least Q40
Consensus quality: 179981 bases at least Q30
Consensus quality: 186957 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 193640; sum-of-contigs
Quality coverage: 6.0 in Q20 bases; sum-of-contigs
Quality coverage: 5.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 1002 1101: contig of 1001 bp in length
* 1002 1101: gap of 100 bp
* 1102 2141: contig of 1040 bp in length
* 2142 2241: gap of 100 bp
* 2242 3546: contig of 1305 bp in length
* 3547 3646: gap of 100 bp
* 3647 4739: contig of 1093 bp in length
* 4740 4839: gap of 100 bp
* 4840 5885: contig of 1046 bp in length
* 5886 5985: gap of 100 bp
* 5986 7222: contig of 1237 bp in length
* 7223 7322: gap of 100 bp
* 7323 8677: contig of 1355 bp in length
* 8678 8777: gap of 100 bp
* 8778 10491: contig of 1714 bp in length
* 10492 10591: gap of 100 bp
* 10592 11828: contig of 1237 bp in length

```

```
* 11829 11928: gap of 100 bp
* 11929 13104: contig of 1176 bp in length
* 13105 13204: gap of 100 bp
* 13205 14205: contig of 1001 bp in length
* 14206 14305: gap of 100 bp
* 14306 15582: contig of 1277 bp in length
* 15583 15682: gap of 100 bp
* 15683 16720: contig of 1038 bp in length
* 16721 16820: gap of 100 bp
* 16821 17978: contig of 1158 bp in length
* 17979 18078: gap of 100 bp
* 18079 19152: contig of 1074 bp in length
* 19153 19252: gap of 100 bp
* 19253 20563: contig of 1311 bp in length
* 20564 20663: gap of 100 bp
* 20664 21761: contig of 1098 bp in length
* 21762 21861: gap of 100 bp
* 21862 22902: contig of 1041 bp in length
* 22903 23002: gap of 100 bp
* 23003 24045: contig of 1043 bp in length
* 24046 24145: gap of 100 bp
* 24146 25692: contig of 1547 bp in length
* 25693 25792: gap of 100 bp
* 25793 27574: contig of 1782 bp in length
* 27575 27674: gap of 100 bp
* 27675 29750: contig of 2076 bp in length
* 29751 29850: gap of 100 bp
* 29851 32791: contig of 2941 bp in length
* 32792 32891: gap of 100 bp
* 32892 35428: contig of 2537 bp in length
* 35429 35528: gap of 100 bp
* 35529 41425: contig of 5897 bp in length
* 41426 41525: gap of 100 bp
* 41526 47276: contig of 5751 bp in length
* 47277 47376: gap of 100 bp
* 47377 52698: contig of 5322 bp in length
* 52699 52798: gap of 100 bp
* 52799 62767: contig of 9969 bp in length
* 62768 62867: gap of 100 bp
* 62868 78028: contig of 15161 bp in length
* 78029 78128: gap of 100 bp
* 78129 94604: contig of 16476 bp in length
* 94605 94704: gap of 100 bp
* 94705 112411: contig of 17707 bp in length
* 112412 112511: gap of 100 bp
* 112512 148635: contig of 36124 bp in length
* 148636 148735: gap of 100 bp
* 148736 196840: contig of 48105 bp in length.
FEATURES
source
1. .196840
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-754A4"
/clone_id="RP11-11 Human Male BAC"
1. .1001
/misc_feature /note="assembly_fragment"
1102. .2141
/misc_feature /note="assembly_fragment"
2242. .3546
/misc_feature /note="assembly_fragment"
3647. .4739
/misc_feature /note="assembly_fragment"
4840. .5885
/misc_feature /note="assembly_fragment"
5986. .7222
/misc_feature /note="assembly_fragment"
7323. .8677
/misc_feature /note="assembly_fragment"
8778. .10491
/misc_feature /note="assembly_fragment"
10592. .11828
/misc_feature /note="assembly_fragment"
11929. .13104
```

```
/note="assembly_fragment"
13205. .14205
/misc_feature /note="assembly_fragment"
14306. .15582
/misc_feature /note="assembly_fragment"
15683. .16720
/misc_feature /note="assembly_fragment"
16821. .17978
/misc_feature /note="assembly_fragment"
18079. .19152
/misc_feature /note="assembly_fragment"
19253. .20563
/misc_feature /note="assembly_fragment"
20664. .21761
/misc_feature /note="assembly_fragment"
21862. .22902
/misc_feature /note="assembly_fragment"
23003. .24045
/misc_feature /note="assembly_fragment"
24146. .25692
/misc_feature /note="assembly_fragment"
25793. .27574
/misc_feature /note="assembly_fragment"
27675. .29750
/misc_feature /note="assembly_fragment"
29851. .32791
/misc_feature /note="assembly_fragment"
32892. .35428
/misc_feature /note="assembly_fragment"
clone_end:T7
vector_side:right"
35529. .41425
/misc_feature /note="assembly_fragment"
41526. .47276
/misc_feature /note="assembly_fragment"
47377. .52698
/misc_feature /note="assembly_fragment"
52799. .62767
/misc_feature /note="assembly_fragment"
62868. .78028
/misc_feature /note="assembly_fragment"
78129. .94604
/misc_feature /note="assembly_fragment"
```

```
Query Match 87.0%; Score 17.4; DB 67; Length 196840;
Best Local Similarity 94.7%; Pred. No. 2e+02; 1; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gagtagaagaagatagaac 19
|||||
Db 43583 GAGTAGAAGATAGAAC 43601
```

```
RESULT 12
AC009245 197663 bp DNA PRI 29-JUN-2000
LOCUS AC009245
DEFINITION Human Chromosome 7 clone RP11-351B12, complete sequence.
ACCESSION AC009245
VERSION AC009245.10 GI:8810492
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 197663)
AUTHORS Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and
Olson,M.V.
TITLE Large-scale Mapping and Sequencing of Human Chromosome 7
JOURNAL 2 (bases 1 to 197663)
REFERENCE Bubb,K.L., Desmarrats,C.L., Ramsey,S.A. and Hubley,R.M.
AUTHORS Direct Submission
TITLE Submitted (07-AUG-1999) Human Genome Center, University of
```

REFERENCE Washington, Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 197663)
 AUTHORS Kaul, R.K., Zhou, Y., James, R.A., Raymond, C., Haugen, E.D. and Olson, M.V.
 TITLE Direct Submission
 JOURNAL Submitted (29-JUN-2000) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
 COMMENT On Jun 29, 2000 this sequence version replaced gi:8039711.
 ----- Genome Center -----
 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: <http://www.genome.washington.edu/UWGC/>
 Contact: uwgchgs@u.washington.edu
 ----- Project Information -----
 Center project name: HaChrt7
 Center clone name: RP11-351B12 (djs263)
 ----- Summary Statistics -----
 Sequencing vector: Plasmid; X52328; 100% of reads
 Chemistry: Dye-primer-amerham; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 1138759 bases at least Q40
 Consensus quality: 1528231 bases at least Q20
 Consensus quality: 1749145 bases at least Q20
 Insert size: 210450; 4.1% error; agarose-1p
 Insert size: 197663; sum-of-contigs
 Quality coverage: 8.31x in Q20 bases; agarose-1p
 Quality coverage: 8.85x in Q20 bases; sum-of-contigs

Overlapping Sequences:
 5': RP11-22814 (UMGC:djs352)
 3': RP11-2983 (UMGC:djs300)

Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. Double stranded (DS) coverage: 92.9%
 DS or two chemistry coverage: 6.82%
 Single stranded regions: 0.258% in 5 gaps

Sequence Validation:
 This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence predicted fragments is given below. The electronically-digested sequence consists of both Insert and vector. In order to accurately represent the entire circular BAC, small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BglII EcoRI
 ----- Ns11 -----

SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt
 SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt

5535	5377	8696	8824	14525	14394
8696	8822				
2067	2046	510	<800	3462	3515
510	<800				
4620	4587	6221	6154	1637	1583
6221	6168				
834	809	12486	12560	286	<800
12486	12543				
10311	10377	97	<800	2229	2327
97	<800				
13707	13646	4074	4231	1264	1215
4074	4202				
1356	1319	12812	12560	4514	4528
12812	12543				
7643	7323	21	<800	705	<800
21	<800				
10251	10377	1282	1331	1944	1909
1282	1336				
1951	2046	1832	1898	1202	1215
1832	1901				
628	<800	4277	4231	119	<800
4277	4202				
3987	3901	912	939	396	<800
912	940				
597	<800	3686	3650	2403	2327
3686	3639				
3588	3552	729	<800	1149	1215
729	<800				
822	809	1339	1331	1576	1583
1339	1336				
5087	5104	1040	1046	2768	2975
1040	1050				
941	919	3112	3135	1442	1417
3112	3138				
5377	5377	964	939	2969	2975
964	940				

```

-----
208      <800      1375      1331      732      <800
1375      1336      -----
136      <800      448      <800      1189      1215
448      <800      -----
784      809      4901      4916      1057      1092
4901      4897      -----
508      <800      466      <800      5203      5174
466      <800      -----
1447      1376      8786      8824      1551      1583
8786      8822      -----
1344      1319      6576      6593      16871      16417
6576      6600      -----
1787      1928      4111      4231      3995      4127
4111      4202      -----
4772      4805      5414      5404      1647      1777
5414      5398      -----
666      <800      2312      2322      1232      1215
2312      2317      -----

```

Query Match 87.0%; Score 17.4; DB 87; Length 197663;
 Best Local Similarity 94.7%; Pred. No. 2e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gagtagaagagtagaac 19
 |||||
 Db 79197 GAGTAGAAGAGTAGAAC 79215

RESULT 13
 AL513347 210859 bp DNA HTG 07-APR-2001
 LOCUS Mus musculus chromosome X clone RP23-476D16, *** SEQUENCING IN
 DEFINITION PROGRESS ***, In unordered pieces.
 ACCESSION AL513347 GI:13561327
 VERSION HTG: HTGS_PHASE1; HTGS_ACTIVERFIN; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS house mouse.
 SOURCE Mus musculus.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 210859)
 Tromans, A.
 Direct Submission
 Submitted (06-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Apr 8, 2001 this sequence version replaced gi:12718187.

 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

 Project Information
 Center project name: Bm476D16

 Summary Statistics

Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; U08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 208726 bases at least Q40
 Consensus quality: 209071 bases at least Q30
 Consensus quality: 209334 bases at least Q20
 Insert size: 209659; sum-of-contigs
 Insert size: 201668; 4.4% error; agarose-fp
 Quality coverage: 13.11x in Q20 bases; sum-of-contigs Quality
 coverage: 13.98x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES

source 1. .210859
 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-476D16"
 /clone_id="RPC1-23"
 1. .4580
 /note="assembly-fragment:05672
 fragment_chain:1"
 4681. .8555
 /note="assembly-fragment:07375
 fragment_chain:1"
 8656. .31018
 /note="assembly-fragment:08111
 fragment_chain:1"
 31119. .34560
 /note="assembly-fragment:02143
 fragment_chain:1"
 34661. .39382
 /note="assembly-fragment:06785
 fragment_chain:1"
 39483. .47527
 /note="assembly-fragment:06336
 fragment_chain:1"
 47628. .72453
 /note="assembly-fragment:07547
 fragment_chain:2"
 72554. .95012
 /note="assembly-fragment:07511
 fragment_chain:2"
 95113. .149922
 /note="assembly-fragment:02768
 fragment_chain:2"
 150023. .160893
 /note="assembly-fragment:00710
 fragment_chain:2"
 160994. .166350
 /note="assembly-fragment:05231
 fragment_chain:2"
 166451. .207988
 /note="assembly-fragment:00818"
 208089. .210859
 /note="assembly-fragment:03090"
 BASE COUNT 63414 a 43413 c 43076 g 59755 t 1201 others
 ORIGIN

Query Match 87.0%; Score 17.4; DB 81; Length 210859;
 Best Local Similarity 94.7%; Pred. No. 2e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gagtagaagagtagaac 19
 |||||
 Db 78060 GAGTAGAAGAGTAGAAC 78078

RESULT 14

AL158210/c
 LOCUS AL158210 191549 bp DNA HTG 08-APR-2001
 DEFINITION Homo sapiens chromosome 10 clone RP11-55418, *** SEQUENCING IN
 PROGRESS ***, 4 unordered pieces.
 ACCESSION AL158210
 VERSION AL158210.10 GI:13567917
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Burton, J.
 TITLE Direct Submission
 JOURNAL Submitted (07-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Apr 9, 2001 this sequence version replaced g1:11228972.
 COMMENT
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: ba55418
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 189400 bases at least Q40
 Consensus quality: 189884 bases at least Q30
 Consensus quality: 190273 bases at least Q20
 Insert size: 191249; sum-of-contrigs
 Insert size: 170480; 7.1% error; agarose-fp
 Quality coverage: 7.73x in Q20 bases; sum-of-contrigs quality
 coverage: 8.72x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 63110: contig of 63110 bp in length
 * 63111 63210: gap of 100 bp
 * 63211 84560: contig of 21350 bp in length
 * 84561 84660: gap of 100 bp
 * 84661 188906: contig of 104246 bp in length
 * 188907 189006: gap of 100 bp
 * 189007 191549: contig of 2343 bp in length.
 Location/Qualifiers
 1. 191549
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-55418"
 /clone_1fb="RPC1-11.2"
 1. 63110
 /note="assembly_fragment:02425
 clone_end:SP6
 vector_side:left"
 63211. 84560
 /note="assembly_fragment:01526
 fragment_chain:1"
 84661. 188906
 /note="assembly_fragment:00575
 fragment_chain:1"
 189007. 191549
 /note="assembly_fragment:00076"
 misc_feature
 misc_feature
 misc_feature
 misc_feature
 BASE COUNT 56923 a 40957 c 38670 g 54690 t 309 others
 ORIGIN

Query Match 85.0%; Score 17; DB 79; Length 191549;
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 88624 GTRGGAAGCTAGCAAC 88608
 3 gtaggaagcataaac 19
 |||||
 RESULT 15
 LOCUS A96152/c 792 bp DNA PAT 07-SEP-2000
 DEFINITION Sequence 185 from Patent WO9924578.
 ACCESSION A96152
 VERSION A96152.1 GI:6779968
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE
 AUTHORS 1 (bases 1 to 792)
 TITLE Pizza, M., Scarlato, V., Rappuoli, R., Grandi, G. and Maignani, V.
 JOURNAL Neisserial antigens
 Patent: WO 9924578-A 185 20-MAY-1999;
 PIZZA MARIAGRAZIA (IT); SCARLATO VINCENZO (IT); RAPPUOLI RINO (IT);
 CHITON SPA (IT); GRANDI GUIDO (IT); MASONANI VEGA (IT)
 Location/Qualifiers
 1. 792
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 210 a 184 c 189 g 209 t
 ORIGIN
 Query Match 84.0%; Score 16.8; DB 9; Length 792;
 Best Local Similarity 90.0%; Pred. No. 2.4e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 148 GACTAGCAATGAAAGCAACG 129
 1 gagtaggaagcataaacg 20
 |||||
 Search completed: October 16, 2001, 19:05:32
 Job time: 4465 sec
